

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
7 September 2001 (07.09.2001)

PCT

(10) International Publication Number
WO 01/64922 A2

(51) International Patent Classification⁷: C12N 15/70,
C07K 14/22, 19/00

Siena (IT). PIZZA, Mariagrazia [IT/IT]; Chiron SpA,
Via Fiorentina, 1, I-53100 Siena (IT).

(21) International Application Number: PCT/IB01/00452

(74) Agents: HALLYBONE, Huw, George et al.; Carpmaels
& Ransford, 43 Bloomsbury Square, London WC1A 2RA
(GB).

(22) International Filing Date: 28 February 2001 (28.02.2001)

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM,
HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK,
LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,
MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL,
TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(25) Filing Language: English

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(26) Publication Language: English

(30) Priority Data:

0004695.3 28 February 2000 (28.02.2000) GB
0027675.8 13 November 2000 (13.11.2000) GB

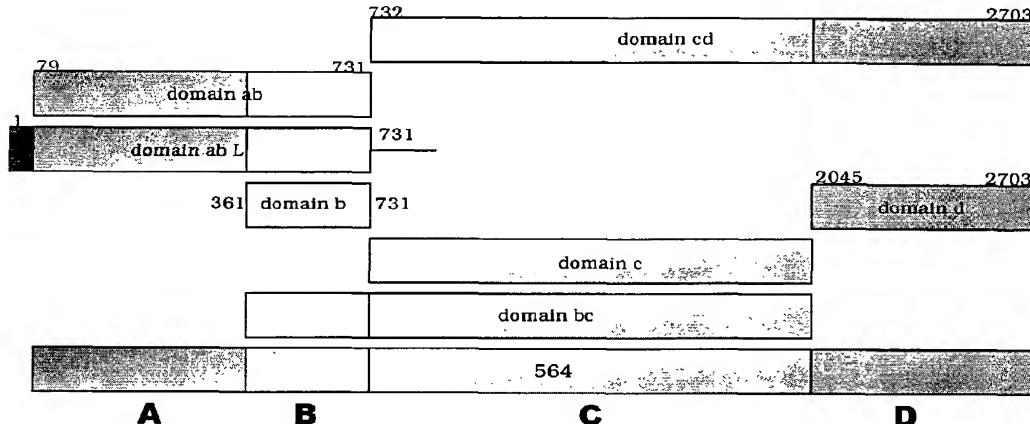
(71) Applicant (for all designated States except US): CHIRON
SPA [IT/IT]; Via Fiorentina 1, I-53100 Siena (IT).

Published:

— without international search report and to be republished
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: HETEROLOGOUS EXPRESSION OF NEISSERIAL PROTEINS



(57) Abstract: Alternative and improved approaches to the heterologous expression of the proteins of *Neisseria meningitidis* and *Neisseria gonorrhoeae*. These approaches typically affect the level of expression, the ease of purification, the cellular localisation, and/or the immunological properties of the expressed protein.

WO 01/64922 A2

HETEROLOGOUS EXPRESSION OF NEISSERIAL PROTEINS

All documents cited herein are incorporated by reference in their entirety.

TECHNICAL FIELD

This invention is in the field of protein expression. In particular, it relates to the heterologous

5 expression of proteins from *Neisseria* (e.g. *N.gonorrhoeae* or, preferably, *N.meningitidis*).

BACKGROUND ART

International patent applications WO99/24578, WO99/36544, WO99/57280 and WO00/22430 disclose proteins from *Neisseria meningitidis* and *Neisseria gonorrhoeae*.

10 These proteins are typically described as being expressed in *E.coli* (i.e. heterologous expression) as either N-terminal GST-fusions or C-terminal His-tag fusions, although other expression systems, including expression in native *Neisseria*, are also disclosed.

15 It is an object of the present invention to provide alternative and improved approaches for the heterologous expression of these proteins. These approaches will typically affect the level of expression, the ease of purification, the cellular localisation of expression, and/or the immunological properties of the expressed protein.

DISCLOSURE OF THE INVENTION

Nomenclature herein

The 2166 protein sequences disclosed in WO99/24578, WO99/36544 and WO99/57280 are referred to herein by the following SEQ# numbers:

Application	Protein sequences	SEQ# herein
WO99/24578	Even SEQ IDs 2-892	SEQ#s 1-446
WO99/36544	Even SEQ IDs 2-90	SEQ#s 447-491
WO99/57280	Even SEQ IDs 2-3020 Even SEQ IDs 3040-3114 SEQ IDs 3115-3241	SEQ#s 492-2001 SEQ#s 2002-2039 SEQ#s 2040-2166

20 In addition to this SEQ# numbering, the naming conventions used in WO99/24578, WO99/36544 and WO99/57280 are also used (e.g. 'ORF4', 'ORF40', 'ORF40-1' etc. as used in WO99/24578 and WO99/36544; 'm919', 'g919' and 'a919' etc. as used in WO99/57280).

The 2160 proteins NMB0001 to NMB2160 from Tettelin *et al.* [Science (2000) 287:1809-1815] are referred to herein as SEQ#s 2167-4326 [see also WO00/66791].

The term 'protein of the invention' as used herein refers to a protein comprising:

- (a) one of sequences SEQ#s 1-4326; or
- 5 (b) a sequence having sequence identity to one of SEQ#s 1-4326; or
- (c) a fragment of one of SEQ#s 1-4326.

The degree of 'sequence identity' referred to in (b) is preferably greater than 50% (e.g. 60%, 70%, 80%, 90%, 95%, 99% or more). This includes mutants and allelic variants [e.g. see WO00/66741]. Identity is preferably determined by the Smith-Waterman homology search 10 algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters *gap open penalty*=12 and *gap extension penalty*=1. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence.

The 'fragment' referred to in (c) should comprise at least *n* consecutive amino acids from 15 one of SEQ#s 1-4326 and, depending on the particular sequence, *n* is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100 or more). Preferably the fragment comprises an epitope from one of SEQ#s 1-4326. Preferred fragments are those disclosed in WO00/71574 and WO01/04316.

Preferred proteins of the invention are found in *N.meningitidis* serogroup B.

20 Preferred proteins for use according to the invention are those of serogroup B *N.meningitidis* strain 2996 or strain 394/98 (a New Zealand strain). Unless otherwise stated, proteins mentioned herein are from *N.meningitidis* strain 2996. It will be appreciated, however, that the invention is not in general limited by strain. References to a particular protein (e.g. '287', '919' etc.) may be taken to include that protein from any strain.

25 ***Non-fusion expression***

In a first approach to heterologous expression, no fusion partner is used, and the native leader peptide (if present) is used. This will typically prevent any 'interference' from fusion partners and may alter cellular localisation and/or post-translational modification and/or folding in the heterologous host.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which (a) no fusion partner is used, and (b) the protein's native leader peptide (if present) is used.

The method will typically involve the step of preparing a vector for expressing a protein of 5 the invention, such that the first expressed amino acid is the first amino acid (methionine) of said protein, and last expressed amino acid is the last amino acid of said protein (*i.e.* the codon preceding the native STOP codon).

This approach is preferably used for the expression of the following proteins using the native leader peptide: 111, 149, 206, 225-1, 235, 247-1, 274, 283, 286, 292, 401, 406, 502-1, 503, 10 519-1, 525-1, 552, 556, 557, 570, 576-1, 580, 583, 664, 759, 907, 913, 920-1, 936-1, 953, 961, 983, 989, Orf4, Orf7-1, Orf9-1, Orf23, Orf25, Orf37, Orf38, Orf40, Orf40.1, Orf40.2, Orf72-1, Orf76-1, Orf85-2, Orf91, Orf97-1, Orf119, Orf143.1, NMB0109 and NMB2050. The suffix 'L' used herein in the name of a protein indicates expression in this manner using the native leader peptide.

15 Proteins which are preferably expressed using this approach using no fusion partner and which have no native leader peptide include: 008, 105, 117-1, 121-1, 122-1, 128-1, 148, 216, 243, 308, 593, 652, 726, 926, 982, Orf83-1 and Orf143-1.

Advantageously, it is used for the expression of ORF25 or ORF40, resulting in a protein which induces better anti-bactericidal antibodies than GST- or His-fusions.

20 This approach is particularly suited for expressing lipoproteins.

Leader-peptide substitution

In a second approach to heterologous expression, the native leader peptide of a protein of the invention is replaced by that of a different protein. In addition, it is preferred that no fusion partner is used. Whilst using a protein's own leader peptide in heterologous hosts can often 25 localise the protein to its 'natural' cellular location, in some cases the leader sequence is not efficiently recognised by the heterologous host. In such cases, a leader peptide known to drive protein targeting efficiently can be used instead.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which (a) the protein's leader peptide is replaced by the leader peptide from a 30 different protein and, optionally, (b) no fusion partner is used.

The method will typically involve the steps of: obtaining nucleic acid encoding a protein of the invention; manipulating said nucleic acid to remove nucleotides that encode the protein's leader peptide and to introduce nucleotides that encode a different protein's leader peptide.

5 The resulting nucleic acid may be inserted into an expression vector, or may already be part of an expression vector. The expressed protein will consist of the replacement leader peptide at the N-terminus, followed by the protein of the invention minus its leader peptide.

The leader peptide is preferably from another protein of the invention (e.g. one of SEQ#s 1-4326), but may also be from an *E.coli* protein (e.g. the OmpA leader peptide) or an *Erwinia carotovora* protein (e.g. the PelB leader peptide), for instance.

10 A particularly useful replacement leader peptide is that of ORF4. This leader is able to direct lipidation in *E.coli*, improving cellular localisation, and is particularly useful for the expression of proteins 287, 919 and ΔG287. The leader peptide and N-terminal domains of 961 are also particularly useful.

15 Another useful replacement leader peptide is that of *E.coli* OmpA. This leader is able to direct membrane localisation of *E.coli*. It is particularly advantageous for the expression of ORF1, resulting in a protein which induces better anti-bactericidal antibodies than both fusions and protein expressed from its own leader peptide.

20 Another useful replacement leader peptide is MKKYLFSAA. This can direct secretion into culture medium, and is extremely short and active. The use of this leader peptide is not restricted to the expression of Neisserial proteins – it may be used to direct the expression of any protein (particularly bacterial proteins).

Leader-peptide deletion

In a third approach to heterologous expression, the native leader peptide of a protein of the invention is deleted. In addition, it is preferred that no fusion partner is used.

25 Thus the invention provides a method for the heterologous expression of a protein of the invention, in which (a) the protein's leader peptide is deleted and, optionally, (b) no fusion partner is used.

30 The method will typically involve the steps of: obtaining nucleic acid encoding a protein of the invention; manipulating said nucleic acid to remove nucleotides that encode the protein's leader peptide. The resulting nucleic acid may be inserted into an expression vector, or may

already be part of an expression vector. The first amino acid of the expressed protein will be that of the mature native protein.

This method can increase the levels of expression. For protein 919, for example, expression levels in *E.coli* are much higher when the leader peptide is deleted. Increased expression

5 may be due to altered localisation in the absence of the leader peptide.

The method is preferably used for the expression of 919, ORF46, 961, 050-1, 760 and 287.

Domain-based expression

In a fourth approach to heterologous expression, the protein is expressed as domains. This may be used in association with fusion systems (*e.g.* GST or His-tag fusions).

10 Thus the invention provides a method for the heterologous expression of a protein of the invention, in which (a) at least one domain in the protein is deleted and, optionally, (b) no fusion partner is used.

15 The method will typically involve the steps of: obtaining nucleic acid encoding a protein of the invention; manipulating said nucleic acid to remove at least one domain from within the protein. The resulting nucleic acid may be inserted into an expression vector, or may already be part of an expression vector. Where no fusion partners are used, the first amino acid of the expressed protein will be that of a domain of the protein.

20 A protein is typically divided into notional domains by aligning it with known sequences in databases and then determining regions of the protein which show different alignment patterns from each other.

25 The method is preferably used for the expression of protein 287. This protein can be notionally split into three domains, referred to as A B & C (see Figure 5). Domain B aligns strongly with IgA proteases, domain C aligns strongly with transferrin-binding proteins, and domain A shows no strong alignment with database sequences. An alignment of polymorphic forms of 287 is disclosed in WO00/66741.

Once a protein has been divided into domains, these can be (a) expressed singly (b) deleted from with the protein *e.g.* protein ABCD → ABD, ACD, BCD *etc.* or (c) rearranged *e.g.* protein ABC → ACB, CAB *etc.* These three strategies can be combined with fusion partners is desired.

ORF46 has also been notionally split into two domains – a first domain (amino acids 1-433) which is well-conserved between species and serogroups, and a second domain (amino acids 433-608) which is not well-conserved. The second domain is preferably deleted. An alignment of polymorphic forms of ORF46 is disclosed in WO00/66741.

5 Protein 564 has also been split into domains (Figure 8), as have protein 961 (Figure 12) and protein 502 (amino acids 28-167 of the MC58 protein).

Hybrid proteins

In a fifth approach to heterologous expression, two or more (*e.g.* 3, 4, 5, 6 or more) proteins of the invention are expressed as a single hybrid protein. It is preferred that no 10 non-Neisserial fusion partner (*e.g.* GST or poly-His) is used.

This offers two advantages. Firstly, a protein that may be unstable or poorly expressed on its own can be assisted by adding a suitable hybrid partner that overcomes the problem. Secondly, commercial manufacture is simplified – only one expression and purification need be employed in order to produce two separately-useful proteins.

15 Thus the invention provides a method for the simultaneous heterologous expression of two or more proteins of the invention, in which said two or more proteins of the invention are fused (*i.e.* they are translated as a single polypeptide chain).

The method will typically involve the steps of: obtaining a first nucleic acid encoding a first protein of the invention; obtaining a second nucleic acid encoding a second protein of the 20 invention; ligating the first and second nucleic acids. The resulting nucleic acid may be inserted into an expression vector, or may already be part of an expression vector.

Preferably, the constituent proteins in a hybrid protein according to the invention will be from the same strain.

The fused proteins in the hybrid may be joined directly, or may be joined via a linker peptide 25 *e.g.* via a poly-glycine linker (*i.e.* G_n where n = 3, 4, 5, 6, 7, 8, 9, 10 or more) or via a short peptide sequence which facilitates cloning. It is evidently preferred not to join a ΔG protein to the C-terminus of a poly-glycine linker.

The fused proteins may lack native leader peptides or may include the leader peptide sequence of the N-terminal fusion partner.

The method is well suited to the expression of proteins orf1, orf4, orf25, orf40, Orf46/46.1, orf83, 233, 287, 292L, 564, 687, 741, 907, 919, 953, 961 and 983.

The 42 hybrids indicated by 'X' in the following table of form NH₂-A—B-COOH are preferred:

↓A	B→	ORF46.1	287	741	919	953	961	983
ORF46.1			X	X	X	X	X	X
287	X			X	X	X	X	X
741	X	X			X	X	X	X
919	X	X	X			X	X	X
953	X	X	X	X			X	X
961	X	X	X	X	X			X
983	X	X	X	X	X	X		

5 Preferred proteins to be expressed as hybrids are thus ORF46.1, 287, 741, 919, 953, 961 and 983. These may be used in their essentially full-length form, or poly-glycine deletions (ΔG) forms may be used (e.g. ΔG -287, ΔG Tbp2, ΔG 741, ΔG 983 etc.), or truncated forms may be used (e.g. $\Delta 1$ -287, $\Delta 2$ -287 etc.), or domain-deleted versions may be used (e.g. 287B, 287C, 287BC, ORF46₁₋₄₃₃, ORF46₄₃₃₋₆₀₈, ORF46, 961c etc.).

10 Particularly preferred are: (a) a hybrid protein comprising 919 and 287; (b) a hybrid protein comprising 953 and 287; (c) a hybrid protein comprising 287 and ORF46.1; (d) a hybrid protein comprising ORF1 and ORF46.1; (e) a hybrid protein comprising 919 and ORF46.1; (f) a hybrid protein comprising ORF46.1 and 919; (g) a hybrid protein comprising ORF46.1, 287 and 919; (h) a hybrid protein comprising 919 and 519; and (i) a hybrid protein comprising ORF97 and 225. Further embodiments are shown in Figure 14.

15

Where 287 is used, it is preferably at the C-terminal end of a hybrid; if it is to be used at the N-terminus, it is preferred to use a ΔG form of 287 is used (e.g. as the N-terminus of a hybrid with ORF46.1, 919, 953 or 961).

Where 287 is used, this is preferably from strain 2996 or from strain 394/98.

20 Where 961 is used, this is preferably at the N-terminus. Domain forms of 961 may be used.

Alignments of polymorphic forms of ORF46, 287, 919 and 953 are disclosed in WO00/66741. Any of these polymorphs can be used according to the present invention.

Temperature

In a sixth approach to heterologous expression, proteins of the invention are expressed at a low temperature.

Expressed Neisserial proteins (*e.g.* 919) may be toxic to *E.coli*, which can be avoided by 5 expressing the toxic protein at a temperature at which its toxic activity is not manifested.

Thus the present invention provides a method for the heterologous expression of a protein of the invention, in which expression of a protein of the invention is carried out at a temperature at which a toxic activity of the protein is not manifested.

A preferred temperature is around 30°C. This is particularly suited to the expression of 919.

10 *Mutations*

As discussed above, expressed Neisserial proteins may be toxic to *E.coli*. This toxicity can be avoided by mutating the protein to reduce or eliminate the toxic activity. In particular, mutations to reduce or eliminate toxic enzymatic activity can be used, preferably using site-directed mutagenesis.

15 In a seventh approach to heterologous expression, therefore, an expressed protein is mutated to reduce or eliminate toxic activity.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which protein is mutated to reduce or eliminate toxic activity.

20 The method is preferably used for the expression of protein 907, 919 or 922. A preferred mutation in 907 is at Glu-117 (*e.g.* Glu→Gly); preferred mutations in 919 are at Glu-255 (*e.g.* Glu→Gly) and/or Glu-323 (*e.g.* Glu→Gly); preferred mutations in 922 are at Glu-164 (*e.g.* Glu→Gly), Ser-213 (*e.g.* Ser→Gly) and/or Asn-348 (*e.g.* Asn→Gly).

Alternative vectors

25 In a eighth approach to heterologous expression, an alternative vector used to express the protein. This may be to improve expression yields, for instance, or to utilise plasmids that are already approved for GMP use.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which an alternative vector is used. The alternative vector is preferably pSM214, with no fusion partners. Leader peptides may or may not be included.

This approach is particularly useful for protein 953. Expression and localisation of 953 with its native leader peptide expressed from pSM214 is much better than from the pET vector.

pSM214 may also be used with: Δ G287, Δ 2-287, Δ 3-287, Δ 4-287, Orf46.1, 961L, 961, 961(MC58), 961c, 961c-L, 919, 953 and Δ G287-Orf46.1.

- 5 Another suitable vector is pET-24b (Novagen; uses kanamycin resistance), again using no fusion partners. pET-24b is preferred for use with: Δ G287K, Δ 2-287K, Δ 3-287K, Δ 4-287K, Orf46.1-K, Orf46A-K, 961-K (MC58), 961a-K, 961b-K, 961c-K, 961c-L-K, 961d-K, Δ G287-919-K, Δ G287-Orf46.1-K and Δ G287-961-K.

Multimeric form

- 10 In a ninth approach to heterologous expression, a protein is expressed or purified such that it adopts a particular multimeric form.

This approach is particularly suited to protein 953. Purification of one particular multimeric form of 953 (the monomeric form) gives a protein with greater bactericidal activity than other forms (the dimeric form).

- 15 Proteins 287 and 919 may be purified in dimeric forms.

Protein 961 may be purified in a 180kDa oligomeric form (*e.g.* a tetramer).

Lipidation

In a tenth approach to heterologous expression, a protein is expressed as a lipidated protein.

- 20 Thus the invention provides a method for the heterologous expression of a protein of the invention, in which the protein is expressed as a lipidated protein.

This is particularly useful for the expression of 919, 287, ORF4, 406, 576-1, and ORF25. Polymorphic forms of 919, 287 and ORF4 are disclosed in WO00/66741.

The method will typically involve the use of an appropriate leader peptide without using an N-terminal fusion partner.

- 25 ***C-terminal deletions***

In an eleventh approach to heterologous expression, the C-terminus of a protein of the invention is mutated. In addition, it is preferred that no fusion partner is used.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which (a) the protein's C-terminus region is mutated and, optionally, (b) no fusion partner is used.

The method will typically involve the steps of: obtaining nucleic acid encoding a protein of the invention; manipulating said nucleic acid to mutate nucleotides that encode the protein's C-terminus portion. The resulting nucleic acid may be inserted into an expression vector, or may already be part of an expression vector. The first amino acid of the expressed protein will be that of the mature native protein.

The mutation may be a substitution, insertion or, preferably, a deletion.

This method can increase the levels of expression, particularly for proteins 730, ORF29 and ORF46. For protein 730, a C-terminus region of around 65 to around 214 amino acids may be deleted; for ORF46, the C-terminus region of around 175 amino acids may be deleted; for ORF29, the C-terminus may be deleted to leave around 230-370 N-terminal amino acids.

Leader peptide mutation

In a twelfth approach to heterologous expression, the leader peptide of the protein is mutated. This is particularly useful for the expression of protein 919.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which the protein's leader peptide is mutated.

The method will typically involve the steps of: obtaining nucleic acid encoding a protein of the invention; and manipulating said nucleic acid to mutate nucleotides within the leader peptide. The resulting nucleic acid may be inserted into an expression vector, or may already be part of an expression vector.

Poly-glycine deletion

In a thirteenth approach to heterologous expression, poly-glycine stretches in wild-type sequences are mutated. This enhances protein expression.

The poly-glycine stretch has the sequence (Gly)_n, where n≥4 (e.g. 5, 6, 7, 8, 9 or more). This stretch is mutated to disrupt or remove the (Gly)_n. This may be by deletion (e.g. CGGGGS→ CGGGS, CGGS, CGS or CS), by substitution (e.g. CGGGGS→ CGXGGGS, CGXXGS, CGXGXS etc.), and/or by insertion (e.g. CGGGGS→ CGGXGGGS, CGXGGGS, etc.).

This approach is not restricted to Neisserial proteins – it may be used for any protein (particularly bacterial proteins) to enhance heterologous expression. For Neisserial proteins, however, it is particularly suitable for expressing 287, 741, 983 and Tbp2. An alignment of polymorphic forms of 287 is disclosed in WO00/66741.

5 Thus the invention provides a method for the heterologous expression of a protein of the invention, in which (a) a poly-glycine stretch within the protein is mutated.

The method will typically involve the steps of: obtaining nucleic acid encoding a protein of the invention; and manipulating said nucleic acid to mutate nucleotides that encode a poly-glycine stretch within the protein sequence. The resulting nucleic acid may be inserted into 10 an expression vector, or may already be part of an expression vector.

Conversely, the opposite approach (*i.e.* introduction of poly-glycine stretches) can be used to suppress or diminish expression of a given heterologous protein.

Heterologous host

Whilst expression of the proteins of the invention may take place in the native host (*i.e.* the 15 organism in which the protein is expressed in nature), the present invention utilises a heterologous host. The heterologous host may be prokaryotic or eukaryotic. It is preferably *E.coli*, but other suitable hosts include *Bacillus subtilis*, *Vibrio cholerae*, *Salmonella typhi*, *Salmonenna typhimurium*, *Neisseria meningitidis*, *Neisseria gonorrhoeae*, *Neisseria lactamica*, *Neisseria cinerea*, *Mycobateria* (*e.g.* *M.tuberculosis*), yeast etc.

20 ***Vectors etc.***

As well as the methods described above, the invention provides (a) nucleic acid and vectors useful in these methods (b) host cells containing said vectors (c) proteins expressed or expressable by the methods (d) compositions comprising these proteins, which may be suitable as vaccines, for instance, or as diagnostic reagents, or as immunogenic compositions 25 (e) these compositions for use as medicaments (*e.g.* as vaccines) or as diagnostic reagents (f) the use of these compositions in the manufacture of (1) a medicament for treating or preventing infection due to Neisserial bacteria (2) a diagnostic reagent for detecting the presence of Neisserial bacteria or of antibodies raised against Neisserial bacteria, and/or (3) a reagent which can raise antibodies against Neisserial bacteria and (g) a method of treating a

patient, comprising administering to the patient a therapeutically effective amount of these compositions.

Sequences

The invention also provides a protein or a nucleic acid having any of the sequences set out in

5 the following examples. It also provides proteins and nucleic acid having sequence identity to these. As described above, the degree of 'sequence identity' is preferably greater than 50% (eg. 60%, 70%, 80%, 90%, 95%, 99% or more).

Furthermore, the invention provides nucleic acid which can hybridise to the nucleic acid disclosed in the examples, preferably under "high stringency" conditions (eg. 65°C in a

10 0.1xSSC, 0.5% SDS solution).

The invention also provides nucleic acid encoding proteins according to the invention.

It should also be appreciated that the invention provides nucleic acid comprising sequences complementary to those described above (eg. for antisense or probing purposes).

Nucleic acid according to the invention can, of course, be prepared in many ways (eg. by

15 chemical synthesis, from genomic or cDNA libraries, from the organism itself etc.) and can take various forms (eg. single stranded, double stranded, vectors, probes etc.).

In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as those containing modified backbones, and also peptide nucleic acids (PNA) etc.

BRIEF DESCRIPTION OF DRAWINGS

20 Figures 1 and 2 show constructs used to express proteins using heterologous leader peptides.

Figure 3 shows expression data for ORF1, and Figure 4 shows similar data for protein 961.

Figure 5 shows domains of protein 287, and Figures 6 & 7 show deletions within domain A.

Figure 8 shows domains of protein 564.

25 Figure 9 shows the *PhoC* reporter gene driven by the 919 leader peptide, and Figure 10 shows the results obtained using mutants of the leader peptide.

Figure 11 shows insertion mutants of protein 730 (A: 730-C1; B: 730-C2).

Figure 12 shows domains of protein 961.

Figure 13 shows SDS-PAGE of Δ G proteins. Dots show the main recombinant product.

Figure 14 shows 26 hybrid proteins according to the invention.

MODES FOR CARRYING OUT THE INVENTION

Example 1 – 919 and its leader peptide

5 Protein 919 from *N.meningitidis* (serogroup B, strain 2996) has the following sequence:

```

1  MKKYILFRAAL YGIAAAITLAA CQSJKSIQTFP QPDTSVINGP DRPVGIPDPA
51  GTTVGGGGAV YTVPVPHLSLP HWAQDFAKS LQSFRRLGCAN LKNRQGWQDV
101 CAQAFQTPVH SFQAKQFFER YFTPWQVAGN GSLAGTVTGY YEPVPLKGDDR
151 RTAQARFPIY GIPDDFISVP LPAGLRSGKA LVRIRQTGKN SGTIDNTGGT
201 HTADLSRFPI TARTTAIKGR FEGSRLFLPYH TRNQINGGAL DGKAPILGYA
251 EDPVELFFMH IQGSGRLKTP SGKYIRIGYA DKNEHPYVSI GRYMADKGYL
301 KLGQTSMQGI KAYMRQNPQR LAEVLGQNPS YIFFRELAGS SNDGPVGALG
351 TPPLMGEYAGA VDRHYITLGA PLFVATAHPV TRKALNRLIM AQDTGSAIKG
401 AVRVDYFWGY GDEAGELAGK QKTTGYVWQL LPNGMKPEYR P*

```

15 The leader peptide is underlined.

The sequences of 919 from other strains can be found in Figures 7 and 18 of WO00/66741.

Example 2 of WO99/57280 discloses the expression of protein 919 as a His-fusion in *E.coli*.

The protein is a good surface-exposed immunogen.

Three alternative expression strategies were used for 919:

20 1) 919 without its leader peptide (and without the mature N-terminal cysteine) and without any fusion partner ('919^{untagged}):

```

1  QSKSIQTFP QPDTSVINGP DRPVGIPDPA GTTVGGGGAV YTVPVPHLSLP
50  HWAQDFAKS LQSFRRLGCAN LKNRQGWQDV CAQAFQTPVH SFQAKQFFER
100 YFTPWQVAGN GSLAGTVTGY YEPVPLKGDDR RTAQARFPIY GIPDDFISVP
150 LPAGLRSGKA LVRIRQTGKN SGTIDNTGGT HTADLSRFPI TARTTAIKGR
200 FEGSRLFLPYH TRNQINGGAL DGKAPILGYA EDPVELFFMH IQGSGRLKTP
250 SGKYIRIGYA DKNEHPYVSI GRYMADKGYL KLGQTSMQGI KAYMRQNPQR
300 LAEVLGQNPS YIFFRELAGS SNDGPVGALG TPPLMGEYAGA VDRHYITLGA
350 PLFVATAHPV TRKALNRLIM AQDTGSAIKG AVRVDYFWGY GDEAGELAGK
400 QKTTGYVWQL LPNGMKPEYR P*

```

The leader peptide and cysteine were omitted by designing the 5'-end amplification primer downstream from the predicted leader sequence.

2) 919 with its own leader peptide but without any fusion partner ('919L'); and

35 3) 919 with the leader peptide (MKTFFKTLsAAALALILAA) from ORF4 ('919LOrf4').

```

1  MKTFFKTLs AAALALALILAA CQSJKSIQTFP QPDTSVINGP DRPVGIPDPA
50  GTTVGGGGAV YTVPVPHLSLP HWAQDFAKS LQSFRRLGCAN LKNRQGWQDV
100 CAQAFQTPVH SFQAKQFFER YFTPWQVAGN GSLAGTVTGY YEPVPLKGDDR
150 RTAQARFPIY GIPDDFISVP LPAGLRSGKA LVRIRQTGKN SGTIDNTGGT
200 HTADLSRFPI TARTTAIKGR FEGSRLFLPYH TRNQINGGAL DGKAPILGYA
250 EDPVELFFMH IQGSGRLKTP SGKYIRIGYA DKNEHPYVSI GRYMADKGYL
300 KLGQTSMQGI KSYMQRQNPQR LAEVLGQNPS YIFFRELAGS SNDGPVGALG

```

350 TPLMGEYAGA VDRHYITLGA PLFVATAHPV TRKALNRLIM AQDTGSAIKG
400 AVRVDYFWGY GDEAGELAGK QKTTGYVWQL LPNGMKPEYR P*

To make this construct, the entire sequence encoding the ORF4 leader peptide was
5 included in the 5'-primer as a tail (primer 919Lorf4 For). A *Nhe*I restriction site was generated by a double nucleotide change in the sequence coding for the ORF4 leader (no amino acid changes), to allow different genes to be fused to the ORF4 leader peptide sequence. A stop codon was included in all the 3'-end primer sequences.

All three forms of the protein were expressed and could be purified.

10 The '919L' and '919LOrf4' expression products were both lipidated, as shown by the incorporation of [³H]-palmitate label. 919^{untagged} did not incorporate the ³H label and was located intracellularly.

15 919LOrf4 could be purified more easily than 919L. It was purified and used to immunise mice. The resulting sera gave excellent results in FACS and ELISA tests, and also in the bactericidal assay. The lipoprotein was shown to be localised in the outer membrane.

919^{untagged} gave excellent ELISA titres and high serum bactericidal activity. FACS confirmed its cell surface location.

Example 2 – 919 and expression temperature

20 Growth of *E.coli* expressing the 919LOrf4 protein at 37°C resulted in lysis of the bacteria. In order to overcome this problem, the recombinant bacteria were grown at 30°C. Lysis was prevented without preventing expression.

Example 3 – mutation of 907, 919 and 922

25 It was hypothesised that proteins 907, 919 and 922 are murein hydrolases, and more particularly lytic transglycosylases. Murein hydrolases are located on the outer membrane and participate in the degradation of peptidoglycan.

The purified proteins 919^{untagged}, 919Lorf4, 919-His (*i.e.* with a C-terminus His-tag) and 922-His were thus tested for murein hydrolase activity [Ursinus & Holtje (1994) *J.Bact.* 176:338-343]. Two different assays were used, one determining the degradation of insoluble murein sacculus into soluble muropeptides and the other measuring breakdown of 30 poly(MurNAc-GlcNAc)_{n>30} glycan strands.

The first assay uses murein sacculi radiolabelled with meso-2,6-diamino-3,4,5-[³H]pimelic acid as substrate. Enzyme (3–10 µg total) was incubated for 45 minutes at 37°C in a total volume of 100µl comprising 10mM Tris-maleate (pH 5.5), 10mM MgCl₂, 0.2% v/v Triton X-100 and [³H]A₂pm labelled murein sacculi (about 10000cpm). The assay mixture was

5 placed on ice for 15 minutes with 100 µl of 1% w/v N-acetyl-N,N,N-trimethylammonium for 15 minutes and precipitated material pelleted by centrifugation at 10000g for 15 minutes. The radioactivity in the supernatant was measured by liquid scintillation counting. *E.coli* soluble lytic transglycosylase SIt70 was used as a positive control for the assay; the negative control comprised the above assay solution without enzyme.

10 All proteins except 919-His gave positive results in the first assay.

The second assay monitors the hydrolysis of poly(MurNAc-GlcNAc)glycan strands. Purified strands, poly(MurNAc-GlcNAc)_{n>30} labelled with N-acetyl-D-1-[³H]glucosamine were incubated with 3µg of 919L in 10 mM Tris-maleate (pH 5.5), 10 mM MgCl₂ and 0.2% v/v Triton X-100 for 30 min at 37°C. The reaction was stopped by boiling for 5 minutes and the 15 pH of the sample adjusted to about 3.5 by addition of 10µl of 20% v/v phosphoric acid. Substrate and product were separated by reversed phase HPLC on a Nucleosil 300 C₁₈ column as described by Harz *et. al.* [Anal. Biochem. (1990) 190:120-128]. The *E.coli* lytic transglycosylase Mlt A was used as a positive control in the assay. The negative control was performed in the absence of enzyme.

20 By this assay, the ability of 919LOrf4 to hydrolyse isolated glycan strands was demonstrated when anhydrodisaccharide subunits were separated from the oligosaccharide by HPLC.

Protein 919Lorf4 was chosen for kinetic analyses. The activity of 919Lorf4 was enhanced 3.7-fold by the addition of 0.2% v/v Triton X-100 in the assay buffer. The presence of Triton X-100 had no effect on the activity of 919^{untagged}. The effect of pH on enzyme activity was 25 determined in Tris-Maleate buffer over a range of 5.0 to 8.0. The optimal pH for the reaction was determined to be 5.5. Over the temperature range 18°C to 42°C, maximum activity was observed at 37°C. The effect of various ions on murein hydrolase activity was determined by performing the reaction in the presence of a variety of ions at a final concentration of 10mM. Maximum activity was found with Mg²⁺, which stimulated activity 2.1-fold. Mn²⁺ and Ca²⁺ 30 also stimulated enzyme activity to a similar extent while the addition Ni²⁺ and EDTA had no significant effect. In contrast, both Fe²⁺ and Zn²⁺ significantly inhibited enzyme activity.

The structures of the reaction products resulting from the digestion of unlabelled *E.coli* murein sacculus were analysed by reversed-phase HPLC as described by Glauner [*Anal. Biochem.* (1988) 172:451-464]. Murein sacculi digested with the muramidase Cellosyl were used to calibrate and standardise the Hypersil ODS column. The major reaction products 5 were 1,6 anhydrodisaccharide tetra and tri peptides, demonstrating the formation of 1,6 anhydromuramic acid intramolecular bond.

These results demonstrate experimentally that 919 is a murein hydrolase and in particular a member of the lytic transglycosylase family of enzymes. Furthermore the ability of 922-His to hydrolyse murein sacculi suggests this protein is also a lytic transglycosylase.

10 This activity may help to explain the toxic effects of 919 when expressed in *E.coli*.

In order to eliminate the enzymatic activity, rational mutagenesis was used. 907, 919 and 922 show fairly low homology to three membrane-bound lipidated murein lytic transglycosylases from *E.coli*:

919 (441aa) is 27.3% identical over 440aa overlap to *E.coli* MLTA (P46885);

15 922 (369aa) is 38.7% identical over 310aa overlap to *E.coli* MLTB (P41052); and

907-2 (207aa) is 26.8% identical over 149aa overlap to *E.coli* MLTC (P52066).

907-2 also shares homology with *E.coli* MLTD (P23931) and Slt70 (P03810), a soluble lytic transglycosylase that is located in the periplasmic space. No significant sequence homology can be detected among 919, 922 and 907-2, and the same is true among the corresponding 20 MLTA, MLTB and MLTC proteins.

Crystal structures are available for Slt70 [1QTEA; 1QTEB; Thunnissen *et al.* (1995) *Biochemistry* 34:12729-12737] and for Slt35 [1LTM; 1QUS; 1QUT; van Asselt *et al.* (1999) *Structure Fold Des* 7:1167-80] which is a soluble form of the 40kDa MLTB.

The catalytic residue (a glutamic acid) has been identified for both Slt70 and MLTB.

25 In the case of Slt70, mutagenesis studies have demonstrated that even a conservative substitution of the catalytic Glu505 with a glutamine (Gln) causes the complete loss of enzymatic activity. Although Slt35 has no obvious sequence similarity to Slt70, their catalytic domains shows a surprising similarity. The corresponding catalytic residue in MLTB is Glu162.

Another residue which is believed to play an important role in the correct folding of the enzymatic cleft is a well-conserved glycine (Gly) downstream of the glutamic acid. Recently, Terrak *et al.* [Mol. Microbiol. (1999) 34:350-64] have suggested the presence of another important residue which is an aromatic amino acid located around 70-75 residues downstream of the catalytic glutamic acid.

Sequence alignment of Slt70 with 907-2 and of MLTB with 922 were performed in order to identify the corresponding catalytic residues in the MenB antigens.

The two alignments in the region of the catalytic domain are reported below:

907-2/S1t70:

922/MLTB

		150	160	▼	170	180	190	200
20	922.pep	VAQKYGVPAELIVAVIGIETNY	<u>GKNT</u>		GFSRVADALATLGFDY	PRRAGFFQKELVELLKLA		
		⋮	⋮	⋮	⋮	⋮	⋮	⋮
	mltb_ecoli	AWQVYGVPP	EII	IVG	I	GVE	TRWGRV	MGKTR
		150	160	▲	170	180	190	200
		GLU162						
25	922.pep	210	220	230	240	250	260	
	mltb_ecoli	KEEGGDVFAFKGSYAGAMGMPQFMPSSYRKWAVD	RDIWGNVGDVAASVAN	YMKQ				
		⋮	⋮	⋮	⋮	⋮	⋮	⋮
		RDEQDDPLNLKGSFAGAMGYQFMPSSYKQYAVD	SGDGHINLWDPV	-DAIGSVAN	YFKA			
		210	220	230	240	250	260	

From these alignments, it results that the corresponding catalytic glutamate in 907-2 is Glu117, whereas in 922 is Glu164. Both antigens also share downstream glycines that could have a structural role in the folding of the enzymatic cleft (in bold), and 922 has a conserved aromatic residue around 70aa downstream (in bold).

35 In the case of protein 919, no 3D structure is available for its *E.coli* homologue MLTA, and nothing is known about a possible catalytic residue. Nevertheless, three amino acids in 919 are predicted as catalytic residues by alignment with MLTA:

919/MLTA

40	240	250	▼	260	□ □	270	□	280	290
	919.pep	ALDGKAPILGYAEDPVELFFMHIQGSGLRKTPSGKYIRI-GYADKNEHPYVSIGRYMADK							
		: : ::: :: : : : : : : : : : : :							
	mlta_ecoli.p	ALSDKY-ILAYSNSLMDNFIMDVQGSGYIDFGDGSPLNFFSYAGKNGHAYRSIGKVLIDR							
			170	180	190	200	210		

-18-

		300	310	320	▼	330	□	340	◊350	◊
919.pep	GYLKLGQTSMQGIKSYMRQNPQ-RLAEVLGQNPSSYIFFRELAGSSNDGPV-GALGTPPLMG	: : :								
5	mlta_ecoli.p	GEVKKEDMSMQAIRHWGETHSEAEVRELLEQNPSFVFFKQPSFA---PVKGASAVPLVG	220	230	240	250	260	270		
		360	▼	○	380		390	400	◊◊410	
10	919.pep	EYAGAVDRHYITLGAPLFVATAHPVTRKALN---RLIMAQDTGSAIKGAVRVDYFWGY	: :							
	mlta_ecoli.p	RASVASDRSIIIPPGTLLAEVPLLDNNNGKFNGQYELRLMVALDVCGAIKGQ-HFDIYQGI	280	290	300	310	320	330		
		420	○							
15	919.pep	GDEAGELAGKQKTTGYVWQLLP	: :							
	mlta_ecoli.p	GPEAGHРАGWYNHYGRVWVLKT	340	350						

The three possible catalytic residues are shown by the symbol ▼:

20 1) Glu255 (Asp in MLTA), followed by three conserved glycines (Gly263, Gly265 and Gly272) and three conserved aromatic residues located approximately 75-77 residues downstream. These downstream residues are shown by □.

25 2) Glu323 (conserved in MLTA), followed by 2 conserved glycines (Gly347 and Gly355) and two conserved aromatic residues located 84-85 residues downstream (Tyr406 or Phe407). These downstream residues are shown by ◊.

3) Asp362 (instead of the expected Glu), followed by one glycine (Gly 369) and a conserved aromatic residue (Trp428). These downstream residues are shown by ○.

Alignments of polymorphic forms of 919 are disclosed in WO00/66741.

Based on the prediction of catalytic residues, three mutants of the 919 and one mutant of 907, containing each a single amino acid substitution, have been generated. The glutamic acids in position 255 and 323 and the aspartic acids in position 362 of the 919 protein and the glutamic acid in position 117 of the 907 protein, were replaced with glycine residues using PCR-based SDM. To do this, internal primers containing a codon change from Glu or Asp to Gly were designed:

Primers	Sequences	Codon change
919-E255 for 919-E255 rev	CGAAGACCCCGTCG <u>gt</u> CTTTTTTTATG GTGCATAAAAAAAAGacCGACGGGGTCT	GAA → Ggt
919-E323 for 919-E323 rev	AACGCCTCGCC <u>Ggt</u> GTTTGGGTCA TTTGACCCAAAAACacCGGCGAGGCG	GAA → Ggt
919-D362 for 919-D362 rev	TGCCGGCGCAGTC <u>Ggt</u> CGGCACTACA TAATGTAGTGCCGacCGACTGCGCCG	GAC → Ggt
907-E117 for 907-E117 rev	TGATTGAGGT <u>GGgt</u> AGCGCGTTCCG GGCGGAACCGCGCTacCCACCTCAAT	GAA → Ggt

Underlined nucleotides code for glycine; the mutated nucleotides are in lower case.

To generate the 919-E255, 919-E323 and 919-E362 mutants, PCR was performed using 20ng of the pET 919-LOrf4 DNA as template, and the following primer pairs:

- 1) Orf4L for / 919-E255 rev
- 2) 919-E255 for / 919L rev
- 3) Orf4L for / 919-E323 rev
- 4) 919-E323 for / 919L rev
- 5) Orf4L for / 919-D362 rev
- 6) 919-D362 for / 919L rev

10 The second round of PCR was performed using the product of PCR 1-2, 3-4 or 5-6 as template, and as forward and reverse primers the "Orf4L for" and "919L rev" respectively.

For the mutant 907-E117, PCR have been performed using 200ng of chromosomal DNA of the 2996 strain as template and the following primer pairs:

- 7) 907L for / 907-E117 rev
- 8) 907-E117 for / 907L rev

The second round of PCR was performed using the products of PCR 7 and 8 as templates and the oligos "907L for" and "907L rev" as primers.

15 The PCR fragments containing each mutation were processed following the standard procedure, digested with *Nde*I and *Xho*I restriction enzymes and cloned into pET-21b+ vector. The presence of each mutation was confirmed by sequence analysis.

Mutation of Glu117 to Gly in 907 is carried out similarly, as is mutation of residues Glu164, Ser213 and Asn348 in 922.

The E255G mutant of 919 shows a 50% reduction in activity; the E323G mutant shows a 70% reduction in activity; the E362G mutant shows no reduction in activity.

Example 4 – multimeric form

287-GST, 919^{untagged} and 953-His were subjected to gel filtration for analysis of quaternary 5 structure or preparative purposes. The molecular weight of the native proteins was estimated using either FPLC Superose 12 (H/R 10/30) or Superdex 75 gel filtration columns (Pharmacia). The buffers used for chromatography for 287, 919 and 953 were 50 mM Tris-HCl (pH 8.0), 20 mM Bicine (pH 8.5) and 50 mM Bicine (pH 8.0), respectively.

10 Additionally each buffer contained 150-200 mM NaCl and 10% v/v glycerol. Proteins were dialysed against the appropriate buffer and applied in a volume of 200µl. Gel filtration was performed with a flow rate of 0.5 – 2.0 ml/min and the eluate monitored at 280nm. Fractions were collected and analysed by SDS-PAGE. Blue dextran 2000 and the molecular weight standards ribonuclease A, chymotrypsin A ovalbumin, albumin (Pharmacia) were used to 15 calibrate the column. The molecular weight of the sample was estimated from a calibration curve of K_{av} vs. $\log M_r$ of the standards. Before gel filtration, 287-GST was digested with thrombin to cleave the GST moiety.

The estimated molecular weights for 287, 919 and 953-His were 73 kDa, 47 kDa and 43 kDa respectively. These results suggest 919 is monomeric while both 287 and 953 are principally dimeric in their nature. In the case of 953-His, two peaks were observed during gel filtration. 20 The major peak (80%) represented a dimeric conformation of 953 while the minor peak (20%) had the expected size of a monomer. The monomeric form of 953 was found to have greater bactericidal activity than the dimer.

Example 5 – pSM214 and pET-24b vectors

953 protein with its native leader peptide and no fusion partners was expressed from the pET 25 vector and also from pSM214 [Velati Bellini *et al.* (1991) *J. Biotechnol.* 18, 177-192].

The 953 sequence was cloned as a full-length gene into pSM214 using the *E. coli* MM294-1 strain as a host. To do this, the entire DNA sequence of the 953 gene (from ATG to the STOP codon) was amplified by PCR using the following primers:

953L for/2 CCGGAATTCTTATGAAAAAAATCATCTTCGCCGC Eco RI

30 953L rev/2 GCCCAAGCTTTATTGTTGGCTGCCTCGATT Hind III

which contain *Eco*RI and *Hind*III restriction sites, respectively. The amplified fragment was digested with *Eco*RI and *Hind*III and ligated with the pSM214 vector digested with the same two enzymes. The ligated plasmid was transformed into *E.coli* MM294-1 cells (by incubation in ice for 65 minutes at 37° C) and bacterial cells plated on LB agar containing 5 20µg/ml of chloramphenicol.

Recombinant colonies were grown over-night at 37°C in 4 ml of LB broth containing 20 µg/ml of chloramphenicol; bacterial cells were centrifuged and plasmid DNA extracted as and analysed by restriction with *Eco*RI and *Hind*III. To analyse the ability of the recombinant colonies to express the protein, they were inoculated in LB broth containing 10 20µg/ml of chloramphenicol and let to grow for 16 hours at 37°C. Bacterial cells were centrifuged and resuspended in PBS. Expression of the protein was analysed by SDS-PAGE and Coomassie Blue staining.

Expression levels were unexpectedly high from the pSM214 plasmid.

Oligos used to clone sequences into pSM-214 vectors were as follows:

ΔG287 (pSM-214)	Fwd	CCGGAATTCTTATG-TCGCCCCATGTTAAATCGGCGGA	EcoRI
	Rev	GCCCAAGCTT-TCAATCCTGCTCTTTTGCCG	HindIII
Δ2 287 (pSM-214)	Fwd	CCGGAATTCTTATG-AGCCAAGATATGGCGGCAGT	EcoRI
	Rev	GCCCAAGCTT-TCAATCCTGCTCTTTTGCCG	HindIII
Δ3 287 (pSM-214)	Fwd	CCGGAATTCTTATG-TCCGCCGAATCCGCAAATCA	EcoRI
	Rev	GCCCAAGCTT-TCAATCCTGCTCTTTTGCCG	HindIII
Δ4 287 (pSM-214)	Fwd	CCGGAATTCTTATG-GGAAGGGTTGATTGGCTAATG	EcoRI
	Rev	GCCCAAGCTT-TCAATCCTGCTCTTTTGCCG	HindIII
Orf46.1 (pSM-214)	Fwd	CCGGAATTCTTATG-TCAGATTGGCAAACGATTCTT	EcoRI
	Rev	GCCCAAGCTT-TTACGTATCATATTTCACGTGCTTC	HindIII
ΔG287-Orf46.1 (pSM-214)	Fwd	CCGGAATTCTTATG-TCGCCCCATGTTAAATCGGCGGA	EcoRI
	Rev	GCCCAAGCTT-TTACGTATCATATTTCACGTGCTTC	HindIII
919 (pSM-214)	Fwd	CCGGAATTCTTATG-CAAAGCAAGAGCATCCAAACCT	EcoRI
	Rev	GCCCAAGCTT-TTACGGCGGTATTGGCT	HindIII
961L (pSM-214)	Fwd	CCGGAATTCATATG-AAACACTTCCATCC	EcoRI
	Rev	GCCCAAGCTT-TTACCACTCGTAATTGAC	HindIII
961 (pSM-214)	Fwd	CCGGAATTCATATG-GCCACAAGCGACGAC	EcoRI
	Rev	GCCCAAGCTT-TTACCACTCGTAATTGAC	HindIII
961c L pSM-214	Fwd	CCGGAATTCTTATG-AAACACTTCCATCC	EcoRI
	Rev	GCCCAAGCTT-TCAACCCACGTTGTAAGGTTG	HindIII
961c pSM-214	Fwd	CCGGAATTCTTATG-GCCACAAACGACGACG	EcoRI
	Rev	GCCCAAGCTT-TCAACCCACGTTGTAAGGTTG	HindIII
953 (pSM-214)	Fwd	CCGGAATTCTTATG-GCCACCTACAAAGTGGACGA	EcoRI
	Rev	GCCCAAGCTT-TTATTGTTGGCTGCCTCGATT	HindIII

These sequences were manipulated, cloned and expressed as described for 953L.

For the pET-24 vector, sequences were cloned and the proteins expressed in pET-24 as described below for pET21. pET2 has the same sequence as pET-21, but with the kanamycin resistance cassette instead of ampicillin cassette.

5 Oligonucleotides used to clone sequences into pET-24b vector were:

ΔG 287 K	Fwd	CGCGGATCC <u>GCTAGC</u> -CCCGATGTTAAATCGGC [§]	NheI
	Rev	CCC <u>GCTCGAG</u> -TCAATCCTGCTTTTTGCC [*]	XhoI
Δ2 287 K	Fwd	CGCGGATCC <u>GCTAGC</u> -CAAGATATGGCGGCAGT [§]	NheI
Δ3 287 K	Fwd	CGCGGATCC <u>GCTAGC</u> -GCCGAATCCGCAAATCA [§]	NheI
Δ4 287 K	Fwd	CG <u>GCTAGC</u> -GGAAGGGTTGATTGGCTAATGG [§]	NheI
Orf46.1 K	Fwd	GGGAATT <u>CCATATG</u> -GGCATTCCCGCAAAATATC	NdeI
	Rev	CCC <u>GCTCGAG</u> -TTACGTATCATATTCACGTGC	XhoI
Orf46A K	Fwd	GGGAATT <u>CCATATG</u> -GGCATTCCCGCAAAATATC	NdeI
	Rev	CCC <u>GCTCGAG</u> -TTATTCTATGCCTTGTGCGGCAT	XhoI
961 K (MC58)	Fwd	CGCGGAT <u>CCCATATG</u> -GCCACAAGCAGCACGA	NdeI
	Rev	CCC <u>GCTCGAG</u> -TTACCACTCGTAATTGAC	XhoI
961a K	Fwd	CGCGGAT <u>CCCATATG</u> -GCCACAAACGACG	NdeI
	Rev	CCC <u>GCTCGAG</u> -TCATTAGCAATATTATCTTGTTC	XhoI
961b K	Fwd	CGCGGAT <u>CCCATATG</u> -AAAGCAAACAGTGCCGAC	NdeI
	Rev	CCC <u>GCTCGAG</u> -TTACCACTCGTAATTGAC	XhoI
961c K	Fwd	CGCGGAT <u>CCCATATG</u> -GCCACAAACGACG	NdeI
	Rev	CCC <u>GCTCGAG</u> -TTAACCCACGTTGTAAGGT	XhoI
961cL K	Fwd	CGCGGAT <u>CCCATATG</u> -ATGAAACACITTCATCC	NdeI
	Rev	CCC <u>GCTCGAG</u> -TTAACCCACGTTGTAAGGT	XhoI
961d K	Fwd	CGCGGAT <u>CCCATATG</u> -GCCACAAACGACG	NdeI
	Rev	CCC <u>GCTCGAG</u> -TCAGTCTGACACTGTTTATCC	XhoI
ΔG 287-919 K	Fwd	CGCGGAT <u>CCGCTAGC</u> -CCCGATGTTAAATCGGC	NheI
	Rev	CCC <u>GCTCGAG</u> -TTACGGCGGTATTGG	XhoI
ΔG 287-Orf46.1 K	Fwd	CGCGGAT <u>CCGCTAGC</u> -CCCGATGTTAAATCGGC	NheI
	Rev	CCC <u>GCTCGAG</u> -TTACGTATCATATTCACGTGC	XhoI
ΔG 287-961 K	Fwd	CGCGGAT <u>CCGCTAGC</u> -CCCGATGTTAAATCGGC	NheI
	Rev	CCC <u>GCTCGAG</u> -TTACCACTCGTAATTGAC	XhoI

* This primer was used as a Reverse primer for all the 287 forms.

[§] Forward primers used in combination with the ΔG278 K reverse primer.

Example 6 – ORF1 and its leader peptide

10 ORF1 from *N.meningitidis* (serogroup B, strain MC58) is predicted to be an outer membrane or secreted protein. It has the following sequence:

1 MKTTDKRTTE THRKAPKTGR IRFSPAYLAI CLSGFILPQA WAGHTYFGIN

5 51 YQYYRDFAEN KGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVSRRNG
 101 101 VAAALVGDQYI VSVAHNGGYN NVDFGAEGRN PDQHRFTYKI VKRNNYKAGT
 151 151 KGHPYGGDYH MPRLIHKFVTD AEPVEMTSYD DGRKYIDQNN YPDRVRIGAG
 201 201 RQYWRSEDE PNNRESSYHI ASAYSWLVGG NTFAQNGSGG GTVNLGSEKI
 251 251 KHSPTYGFLPT GGSFGDSGSP MFIYDAQKQWLINGVILQTG NPYIGKSNGF
 301 301 QLVRKDWFYD EIFAGDTHSV FYEPRQNGKY SFNDDNNNGTG KINAKHEHNS
 351 351 LPNRLKTRTV QLFNVSLSET AREPVYHAAG GVNNSYRPRLN NGENISFIDE
 401 401 GKGEILTTSN INQAGGLYF QGDFTVSPEN NETWQGAGVH ISEDSTVTWK
 451 451 VNGVANDRLS KIGKGTLHVQ AKGENQGSIS VGDGTVILDQ QADDKGKKQA
 501 501 FSEIGLVSQR GTVQLNADNQ FNPDKLYFGF RGGRLDLNQH SLSFHRIQNT
 551 551 DEGAMIVNHN QDKESTVTIT GNKDIATTGN NNSLDSKKEI AYNGWFGEKD
 601 601 TTTKTNGRNLN VYQPAEADRT LLLSGGTNLN GNITQTNGKL FFSGRPTPH
 651 651 YNHLDNDHWSQ KEGIPRGEIV WNDNDWINRTF KAENFOIKGG QAVVSRNVAK
 701 701 VKGDWHLSNH AQAVFGVAPH QSHTICTRSD WTGLTNCVEK TITDDKVIAS
 751 751 LTKTDISGNV DLADAHHLNL TGLATLNGNL SANGDTRYTV SHNATQNGNL
 801 801 SLVGNAQATF NQATLNGNTS ASGNASFNLN DHAQNGSLT LSGNAKANVS
 851 851 HSALNGNVSL ADKAVFHFE SRTFGQISGG KDTALHLKDS EWTLPSGTEL
 901 901 GNLNLDNATI TLNSAYRHDA AGAQQTGSATD APRRRSRRSR RSLLSVTPPT
 951 951 SVESRFNTLT VNGKLNQGQT FRFMSELFGY RSDKLKLAES SEGYTLAVN
 20 1001 1001 NTGNEPASLE QLTVVEGKDN KPLSENLNFT LQNEHVDAGA WRYQLIRKD
 1051 1051 EFRLHNPVKE QELSDKLGKA EAKKQAEKDN AQSLDALIAA GRDAVEKTES
 1101 1101 VAEPARQAGG ENVGIMQAAE EKKRVQADKD TALAKQREAE TRPATTAFPR
 1151 1151 ARRARRDLPQ LQPQPOPQPO RDLISRYANS GLSEFSATLN SVFAVQDELD
 1201 1201 RVFAEDRRNA VWTSGIRDTK HYRSQDFRAY RQQTDLRQIG MQKNLGSGRV
 25 1251 1251 GILFSHNRTE NTFDDGIGNS ARLAHGAVFG QYQIDRFYIG ISAGAGFSSG
 1301 1301 SLSDGIGGK IRRVLYHYGIQ ARYRAGFGGF GIEPHIGATR YFVQKADYRY
 1351 1351 ENVNIATPGL AFNRYRAGIK ADYSFKPAQH ISITPYLSSL YTDAASGKVR
 1401 1401 TRVNTAVLAQ DFGKTRSAEW GVNAEIKGFT LSLHAAAAGK PQLEAQHSAG
 1451 1451 IKLGYRW*

30 The leader peptide is underlined.

A polymorphic form of ORF1 is disclosed in WO99/55873.

Three expression strategies have been used for ORF1:

35 1) ORF1 using a His tag, following WO99/24578 (ORF1-His);
 2) ORF1 with its own leader peptide but without any fusion partner ('ORF1L'); and
 3) ORF1 with the leader peptide (MKKTAIAIAVALAGFATVAQA) from *E.coli* OmpA
 ('Orf1LOmpA'):

40 MKKTAIAIAVALAGFATVAQAASAGHTYFGINYQYYRDFAENKGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSV
 VSRNGVAALVGDQYI VSVAHNGGYN NVDFGAEGRN PDQHRFTYKI VIKRNNYKAGT KGHPYGGDYHMPRLHKFVTDAE
 PVEMTSYMDGRKYIDQNNY PDRVRIGAGRQYWRSEDEDE PNNRESSYHI ASAYSWLVGGNTFAQNGSGGGTVNLGSEK
 IKHSPTYGFLPTGGSGDGSQPMFIYDAQKQWLINGVILQTG NPYIGKSNGF QLVRKDWFYD EIFAGDTHSVFYEPRO
 NGKYSFNDDNNNGTGT KINAKHEHNS PNLRLKTRTVQLFNVSLSET AREPVYHAAGGVNSYRPRLNNGENISFIDEKGK
 ELIITSNINQAGGLYF QGDFTVSPENNETWQGAGVH ISEDSTVTWVNGVANDRLS KIGKTLHVQAKGENQGSIS
 VGDGTVILDQADDKGKKQAFSE IGLVSGRTVQLNADNQFNPDKLYFGFRGGRLDLNQH SLSFHRIQNT DEGAMIV
 NHNQDKESTVTITGNKDIATTGNNNSLDSKKEI AYNGWFGEKD TTTKTNGRNLVY QPAAEDRTLLS SGGTNLNQGNIT
 QTNGKLFFSGRPTPHAYNHLNDHWSQKEGIPRGEIV WNDNDWINRTFKAENFOIKGGQAVSRNVAKVGDWHLNHA
 QAVFGVAPHQSHTICTRS DWTLGNTNCVEK TITDDKVIASLT KTDISGNV DLAHHLNL TGLATLNGNL SANGDTRY
 TVSHNATQNGNLSLVGNAQATFNQATLNGNTS ASGNASFNLSD HAVQNGSLT LSGNAKANVSHSALNGNVSLADKAV
 FHFESSRFTGQISGGKDTALHLKDSEWTLPSGTELGNLNLNDNATITLNSAYRHDAAGAQQTGSATDAPRRRSRRSRRS
 LLSVTPPTSVESRFNTLT VNGKLNQGQTFRFMSELFGYRSDKLKLAES SEGTYTLAVNNTGNEPASLEQLTVVEGKD
 NKPLSENLNFTLQNEHVDAGAWRYQOLIRKDGEFRLLHNPVKEQELSDKLGKA EAKKQAEKDNAQSLDALIAAGRDAVE
 KTESVAEPRQAGGENVGIMQAEEEKRVQADKD TALAKQREAE TRPATTAFPRARRRDL PQLQPQWPQQRD
 ISRYANSGLSEF SATLNSVFAVQDELDRVFAEDRRNAWWTSGIRDTRKHYRSQDFRAY RQQTDLRQIGMQKNLGSGRV
 GILFSHNRTE NTFDDGIGNSARLAHGAVFGQYQIDRFYIG ISAGAGFSSGSLSDGIGGKIRRRVLYHGIQARYRAGF
 CGFGIEPHIGATRYFVQKADYRYENVIATPGLAFNRYRAGIKADYSFKPAQH ISITPYLSSL SYTDAASGKVRTRVN
 55 TAVLAQDFGKTRSAEWGVNAEIKGFT LSLHAAAAGK PQLEAQHSAGIKLGYRW*

To make this construct, the clone pET911LOmpA (see below) was digested with the *Nhe*I and *Xho*I restriction enzymes and the fragment corresponding to the vector carrying the OmpA leader sequence was purified (pETLOmpA). The ORF1 gene coding for the mature protein was amplified using the oligonucleotides ORF1-For and ORF1-Rev (including the *Nhe*I and *Xho*I restriction sites, respectively), digested with *Nhe*I and *Xho*I and ligated to the purified pETOmpA fragment (see Figure 1). An additional AS dipeptide was introduced by the *Nhe*I site.

All three forms of the protein were expressed. The His-tagged protein could be purified and was confirmed as surface exposed, and possibly secreted (see Figure 3). The protein was used to immunise mice, and the resulting sera gave excellent results in the bactericidal assay.

ORF1LOmpA was purified as total membranes, and was localised in both the inner and outer membranes. Unexpectedly, sera raised against ORF1LOmpA show even better ELISA and anti-bactericidal properties than those raised against the His-tagged protein.

ORF1L was purified as outer membranes, where it is localised.

15 ***Example 7 – protein 911 and its leader peptide***

Protein 911 from *N.meningitidis* (serogroup B, strain MC58) has the following sequence:

1 MKKNILEFWV GLFVLIGAAA VAFLAFLRVAG GAAFGGSDKT YAVYADFGDI
 51 GGLIKVNAPVK SAGVLVGRVG AIGLDPKSYQ ARVRLDDGK YQFSSDVSAQ
 101 IILTSGLLGEQ YIGLQQGGDT ENLAAGDTIS VTSSAMVLEN LIGKFMITSFA
 151 EKNADGGNAE KAAE*

The leader peptide is underlined.

Three expression strategies have been used for 911:

- 1) 911 with its own leader peptide but without any fusion partner ('911L');
- 2) 911 with the leader peptide from *E.coli* OmpA ('911LOmpA').

25 To make this construct, the entire sequence encoding the OmpA leader peptide was included in the 5'- primer as a tail (primer 911LOmpA Forward). A *Nhe*I restriction site was inserted between the sequence coding for the OmpA leader peptide and the 911 gene encoding the predicted mature protein (insertion of one amino acid, a serine), to allow the use of this construct to clone different genes downstream the OmpA leader peptide sequence.

- 3) 911 with the leader peptide (MKYLLPTAAAGLLLAAQPAMA) from *Erwinia carotovora* PelB ('911LpelB').

To make this construct, the 5'-end PCR primer was designed downstream from the leader sequence and included the *Nco*I restriction site in order to have the 911 fused directly to the PelB leader sequence; the 3'- end primer included the STOP codon. The expression vector used was pET22b+ (Novagen), which carries the coding sequence for the PelB leader peptide. The *Nco*I site introduces an additional methionine after the PelB sequence.

All three forms of the protein were expressed. ELISA titres were highest using 911L, with 919LOmpA also giving good results.

Example 8 – ORF46

10 The complete ORF46 protein from *N.meningitidis* (serogroup B, strain 2996) has the following sequence:

1	<u>LGISRKISLI</u>	<u>LSILAVCLPM</u>	<u>HAHASDLAND</u>	SFIRQVLDLDRQ	HFEPEPDGKYHL
51	FGSRGELAER	SGHIGLGKIQ	SHQLGNLMIQ	QAAIKGNIGY	IVRFSDHGHE
101	VHSPFDNHAS	HSDSDEAGSP	VDGFSLYRIH	WDGYEHHPAD	GYDGPQGGGY
151	PAPKGARDIY	SYDIKGVAQN	IRNLNTDNRS	TGQRLADRFH	NAGSMLTQGV
201	GDGFKRATRY	SPEDLRSGNA	AEAFNGTADI	VKNIIGAAGE	IVGAGDAVQG
251	ISEGNSNIAVM	HGLGLLSTEN	KMARINDLAD	MAQLKDYAAA	AIRDWAVQNP
301	NAAQGIEAVS	NIFMAAIPIK	GIGAVRGKYG	LGGITAHPIK	RSQMGAIALP
351	KGKSAVSDMF	ADAAYAKYPS	PYHSRNIRSN	LEQRYGKENI	TSSTVPPSNG
401	KNVKLADQRH	PKTGVPFDGK	GFPNFEKHVK	YDTKLDIQEL	SGGGIPKAKP
451	VSDAKPRWEV	DRKLNLKLTTR	EQVEKNVQEI	RNGNKNNSNFS	QHAQLEREIN
501	KLKSADEINF	ADGMGKFTDS	MNDKAFLSRLV	KSVKENGFTN	PVVEYVEING
551	KAYIVRGNMR	VFAAEYLGRI	HELKFKKVDF	PVPNTSWKNP	TDVILNESGNV
601	KRPRYRSK*				

25

The leader peptide is underlined.

The sequences of ORF46 from other strains can be found in WO00/66741.

Three expression strategies have been used for ORF46:

- 1) ORF46 with its own leader peptide but without any fusion partner ('ORF46-2L');
- 30 2) ORF46 without its leader peptide and without any fusion partner ('ORF46-2'), with the leader peptide omitted by designing the 5'-end amplification primer downstream from the predicted leader sequence:

1	SDLANDSFIR	QVLDQHFEPE	DGKYHLFGSR	GELAERSGHI	GLGKIQSHQL
51	GNLMIQQAAI	KGNIGYIVRF	SDHGHEVHSP	FDNHASHSDS	DEAGSPVDGF
101	SLYRIHWGQY	EHHPADGYDG	PQGGGYPAPK	GARDIYSYDI	KGVAQNIRLN
151	LTDNRSTGQR	LADRFHNAGS	MLTQGVGDGF	KRATRYSPEL	DRSGNAAEAF
201	NGTADIVKNI	IGAAGEIVGA	GDAVQGISEG	SNIAVMHGLG	LLSTENKMAR
251	INDLADMAQL	KDYAAAIRD	WAVQNPNAAQ	GIEAVSNIFM	AAIPIKGIGA
301	VRGKYGLGGI	TAHPIKRSQM	GAIALPKGKS	AVSDNFADAA	YAKYPSPYHS
351	RNIRSNEQR	YGKENITSST	VPPSNGKNVK	LADQRHPKTG	VPFDGKGFPN
401	FEKHKVYDTK	LDIQELSGGG	IPKAKPVSDA	KPRWEVDRKL	NKLTTREQVE
451	KNVQEIRNGN	KNSNFSQHAQ	LEREINKLKS	ADEINFADGM	GKFTDSMNDK
501	AFSRLVKSVK	ENGFTNPVVE	YVEIINGKAYI	VRGNNRVFAA	EYLGRIHELK
551	FKKVDFPVPN	TSWKNPNDVL	NESGNVKRPR	YRSK*	

3) ORF46 as a truncated protein, consisting of the first 433 amino acids ('ORF46.1L'), constructed by designing PCR primers to amplify a partial sequence corresponding to aa 1-433.

5 A STOP codon was included in the 3'-end primer sequences.

ORF46-2L is expressed at a very low level to *E.coli*. Removal of its leader peptide (ORF46-2) does not solve this problem. The truncated ORF46.1L form (first 433 amino acids, which are well conserved between serogroups and species), however, is well-expressed and gives excellent results in ELISA test and in the bactericidal assay.

10 ORF46.1 has also been used as the basis of hybrid proteins. It has been fused with 287, 919, and ORF1. The hybrid proteins were generally insoluble, but gave some good ELISA and bactericidal results (against the homologous 2996 strain):

Protein	ELISA	Bactericidal Ab
Orf1-Orf46.1-His	850	256
919-Orf46.1-His	12900	512
919-287-Orf46-His	n.d.	n.d.
Orf46.1-287His	150	8192
Orf46.1-919His	2800	2048
Orf46.1-287-919His	3200	16384

For comparison, 'triple' hybrids of ORF46.1, 287 (either as a GST fusion, or in Δ G287 form) and 919 were constructed and tested against various strains (including the homologous 2996 strain) *versus* a simple mixture of the three antigens. FCA was used as adjuvant:

	2996	BZ232	MC58	NGH38	F6124	BZ133
Mixture	8192	256	512	1024	>2048	>2048
ORF46.1-287-919his	16384	256	4096	8192	8192	8192
ΔG287-919-ORF46.1his	8192	64	4096	8192	8192	16384
ΔG287-ORF46.1-919his	4096	128	256	8192	512	1024

Again, the hybrids show equivalent or superior immunological activity.

Hybrids of two proteins (strain 2996) were compared to the individual proteins against various heterologous strains:

	1000	MC58	F6124 (MenA)
ORF46.1-His	<4	4096	<4
ORF1-His	8	256	128
ORF1—ORF46.1-His	1024	512	1024

Again, the hybrid shows equivalent or superior immunological activity.

Example 9 – protein 961

The complete 961 protein from *N.meningitidis* (serogroup B, strain MC58) has the following sequence:

5 1 MSMKHFPAKV LTTAILLATFC SGALAATSDD DVKKAATVAI VAAYNNNGQEI
 51 NGFKAGETIY DIGEDGTITQ KDATAADVEA DDFKGLGLKK VVTNLTKTVN
 10 101 ENKQNVDAKV KAAESEIEKLT TTKLADTDAA LADTDAAALDE TTNALNKLGE
 151 NITTFAAEETK TNIVKIDEKL EAVADTVVDKH AEAFFNDIADS LDETNTKADE
 201 AVKTANEAKQ TAEETKQNVD AKVKAETAA GKAEEAAAGTA NTAADKAEAV
 251 AAKVTDIKAD IATNKADI AK NSARIDS LDK NVANLRKETR QGLAEQAAALS
 301 GLFQPYNVGR FNVTAAVGGY KSESAVAIGT GFRFTENFAA KAGVAVGTSS
 351 GSSAAYHVGV NYEW*

The leader peptide is underlined.

15 Three approaches to 961 expression were used:

- 1) 961 using a GST fusion, following WO99/57280 ('GST961');
- 2) 961 with its own leader peptide but without any fusion partner ('961L'); and
- 3) 961 without its leader peptide and without any fusion partner ('961^{untagged}'), with the leader peptide omitted by designing the 5'-end PCR primer downstream from the predicted leader sequence.

All three forms of the protein were expressed. The GST-fusion protein could be purified and antibodies against it confirmed that 961 is surface exposed (Figure 4). The protein was used to immunise mice, and the resulting sera gave excellent results in the bactericidal assay. 961L could also be purified and gave very high ELISA titres.

25 Protein 961 appears to be phase variable. Furthermore, it is not found in all strains of *N.meningitidis*.

Example 10 – protein 287

Protein 287 from *N.meningitidis* (serogroup B, strain 2996) has the following sequence:

30 1 MFERSVTIAMA CIFALSACGG GGGGSPDVKS ADTLSKPAAP VVAEKETEVK
 51 EDAPQAGSQG QGAPSTQGSQ DMAAVSAENT GNNGAATTDK PKNEDEGPQN
 101 DMPQNSAESA NQTGNQNQPAD SSDSAPASNP APANGGSNFG RVDLANGVLI
 151 DGPSQNITLT HCKGDSCNGD NLLDEEAPSK SEFENLNESE RIEKYKKDGK

201 SDKFTNLVAT AVQANGTNKY VIIYKDKSAS SSSARFRRSA RSRRSLPAEM
 251 PLIPVNQADT LIVDGEAVSL TGHSGNIFAP EGNYRYLTG AEKLPGGSYA
 301 LRVQGEPAKG EMLAGTAVYN GEVLHFHTEN GRPYPTRGRF AAKVDFGSKS
 351 VDGIIDSGDD LHMGTQKFKA AIDGNGFKGT WTENGGDVS GRFYGPAGEE
 5 401 VAGKYSYRPT DAEKGFGVFA AGKKEQD*

The leader peptide is shown underlined.

The sequences of 287 from other strains can be found in Figures 5 and 15 of WO00/66741.

Example 9 of WO99/57280 discloses the expression of 287 as a GST-fusion in *E.coli*.

10 A number of further approaches to expressing 287 in *E.coli* have been used, including:

- 1) 287 as a His-tagged fusion ('287-His');
- 2) 287 with its own leader peptide but without any fusion partner ('287L');
- 3) 287 with the ORF4 leader peptide and without any fusion partner ('287LOrf4'); and
- 4) 287 without its leader peptide and without any fusion partner ('287^{untagged}');

15 1 CGGGGGGSPD VKSADTLSKP AAPVVAEKET EVKEDAPQAG SQGQGAPSTQ
 51 51 GSQDMAAVSA ENTGNNGAAT TDKPKNEDEG PQNDMPQNSA ESANQTGNNQ
 101 101 PADSSDSAPA SNPAPANGGS NFGRVDSLNG VLIDGPSQNI TLTHCKGDSC
 151 151 NGDNLLDEEA PSKSEFENLN ESERIEKYKK DGKSDKFTNL VATAVQANGT
 20 201 NKVIIYKDK SASSSSARFR RSARSRRSLP AEMPLIPVNQ ADTLIVDGEA
 251 251 VSLTGHSQNI FAPEGNYRYL TYGAEKLPGG SYALRVQGEP AKGEMLAGTA
 301 301 VYNGEVLHFH TENGPRYPTR GRFAAKVDFG SKSVDGIIDS GDDLHMGTQK
 351 351 FKAAIDGNGF KGTWTENGDD DVSGRFYGPAA GEEVAGKSY RPTDAEKGGF
 401 401 GVFAAGKKEQD *

25 All these proteins could be expressed and purified.

'287L' and '287LOrf4' were confirmed as lipoproteins.

As shown in Figure 2, '287LOrf4' was constructed by digesting 919LOrf4 with *Nhe*I and *Xho*I. The entire ORF4 leader peptide was restored by the addition of a DNA sequence coding for the missing amino acids, as a tail, in the 5'-end primer (287LOrf4 for), fused to 30 287 coding sequence. The 287 gene coding for the mature protein was amplified using the oligonucleotides 287LOrf4 For and Rev (including the *Nhe*I and *Xho*I sites, respectively), digested with *Nhe*I and *Xho*I and ligated to the purified pETOrf4 fragment.

Example 11 – further non-fusion proteins with/without native leader peptides

A similar approach was adopted for *E.coli* expression of further proteins from WO99/24578, 35 WO99/36544 and WO99/57280.

The following were expressed without a fusion partner: 008, 105, 117-1, 121-1, 122-1, 128-1, 148, 216, 243, 308, 593, 652, 726, 982, and Orf143-1. Protein 117-1 was confirmed as surface-exposed by FACS and gave high ELISA titres.

The following were expressed with the native leader peptide but without a fusion partner:

5 111, 149, 206, 225-1, 235, 247-1, 274, 283, 286, 292, 401, 406, 502-1, 503, 519-1, 525-1, 552, 556, 557, 570, 576-1, 580, 583, 664, 759, 907, 913, 920-1, 926, 936-1, 953, 961, 983, 989, Orf4, Orf7-1, Orf9-1, Orf23, Orf25, Orf37, Orf38, Orf40, Orf40.1, Orf40.2, Orf72-1, Orf76-1, Orf85-2, Orf91, Orf97-1, Orf119, Orf143.1. These proteins are given the suffix 'L'.

His-tagged protein 760 was expressed with and without its leader peptide. The deletion of 10 the signal peptide greatly increased expression levels. The protein could be purified most easily using 2M urea for solubilisation.

His-tagged protein 264 was well-expressed using its own signal peptide, and the 30kDa protein gave positive Western blot results.

All proteins were successfully expressed.

15 The localisation of 593, 121-1, 128-1, 593, 726, and 982 in the cytoplasm was confirmed.

The localisation of 920-1L, 953L, ORF9-1L, ORF85-2L, ORF97-1L, 570L, 580L and 664L in the periplasm was confirmed.

20 The localisation of ORF40L in the outer membrane, and 008 and 519-1L in the inner membrane was confirmed. ORF25L, ORF4L, 406L, 576-1L were all confirmed as being localised in the membrane.

Protein 206 was found not to be a lipoprotein.

25 ORF25 and ORF40 expressed with their native leader peptides but without fusion partners, and protein 593 expressed without its native leader peptide and without a fusion partner, raised good anti-bactericidal sera. Surprisingly, the forms of ORF25 and ORF40 expressed without fusion partners and using their own leader peptides (*i.e.* 'ORF25L' and 'ORF40L') give better results in the bactericidal assay than the fusion proteins.

Proteins 920L and 953L were subjected to N-terminal sequencing, giving HRVWVETAH and ATYKVDEYHANARFAF, respectively. This sequencing confirms that the predicted leader peptides were cleaved and, when combined with the periplasmic location, confirms that the

proteins are correctly processed and localised by *E.coli* when expressed from their native leader peptides.

The N-terminal sequence of protein 519.1L localised in the inner membrane was MEFFIILLA, indicating that the leader sequence is not cleaved. It may therefore function as both an 5 uncleaved leader sequence and a transmembrane anchor in a manner similar to the leader peptide of PBP1 from *N.gonorrhoeae* [Ropp & Nicholas (1997) *J. Bact.* 179:2783-2787.]. Indeed the N-terminal region exhibits strong hydrophobic character and is predicted by the Tmpred. program to be transmembrane.

Example 12 – lipoproteins

10 The incorporation of palmitate in recombinant lipoproteins was demonstrated by the method of Kraft *et. al.* [J. Bact. (1998) 180:3441-3447.]. Single colonies harbouring the plasmid of interest were grown overnight at 37°C in 20 ml of LB/Amp (100µg/ml) liquid culture. The culture was diluted to an OD₅₅₀ of 0.1 in 5.0 ml of fresh medium LB/Amp medium containing 5 µC/ml [³H] palmitate (Amersham). When the OD₅₅₀ of the culture reached 0.4- 15 0.8, recombinant lipoprotein was induced for 1 hour with IPTG (final concentration 1.0 mM). Bacteria were harvested by centrifugation in a bench top centrifuge at 2700g for 15 min and washed twice with 1.0 ml cold PBS. Cells were resuspended in 120µl of 20 mM Tris-HCl (pH 8.0), 1 mM EDTA, 1.0% w/v SDS and lysed by boiling for 10 min. After centrifugation at 13000g for 10 min the supernatant was collected and proteins precipitated 20 by the addition of 1.2 ml cold acetone and left for 1 hour at -20 °C. Protein was pelleted by centrifugation at 13000g for 10 min and resuspended in 20-50µl (calculated to standardise loading with respect to the final O.D of the culture) of 1.0% w/v SDS. An aliquot of 15 µl was boiled with 5µl of SDS-PAGE sample buffer and analysed by SDS-PAGE. After electrophoresis gels were fixed for 1 hour in 10% v/v acetic acid and soaked for 30 minutes 25 in Amplify solution (Amersham). The gel was vacuum-dried under heat and exposed to Hyperfilm (Kodak) overnight -80 °C.

Incorporation of the [³H] palmitate label, confirming lipidation, was found for the following proteins: Orf4L, Orf25L, 287L, 287LOrf4, 406.L, 576L, 926L, 919L and 919LOrf4.

Example 13 – domains in 287

30 Based on homology of different regions of 287 to proteins that belong to different functional classes, it was split into three ‘domains’, as shown in Figure 5. The second domain shows

homology to IgA proteases, and the third domain shows homology to transferrin-binding proteins.

Each of the three 'domains' shows a different degree of sequence conservation between *N.meningitidis* strains – domain C is 98% identical, domain A is 83% identical, whilst 5 domain B is only 71% identical. Note that protein 287 in strain MC58 is 61 amino acids longer than that of strain 2996. An alignment of the two sequences is shown in Figure 7, and alignments for various strains are disclosed in WO00/66741 (see Figures 5 and 15 therein).

The three domains were expressed individually as C-terminal His-tagged proteins. This was done for the MC58 and 2996 strains, using the following constructs:

10 287a-MC58 (aa 1-202), 287b-MC58 (aa 203-288), 287c-MC58 (aa 311-488).
 287a-2996 (aa 1-139), 287b-2996 (aa 140-225), 287c-2996 (aa 250-427).

To make these constructs, the stop codon sequence was omitted in the 3'-end primer sequence. The 5' primers included the *Nhe*I restriction site, and the 3' primers included a *Xho*I as a tail, in order to direct the cloning of each amplified fragment into the expression 15 vector pET21b+ using *Nde*I-*Xho*I, *Nhe*I-*Xho*I or *Nde*I-*Hind*III restriction sites.

All six constructs could be expressed, but 287b-MC8 required denaturation and refolding for solubilisation.

Deletion of domain A is described below ('Δ4 287-His').

20 Immunological data (serum bactericidal assay) were also obtained using the various domains from strain 2996, against the homologous and heterologous MenB strains, as well as MenA (F6124 strain) and MenC (BZ133 strain):

	2996	BZ232	MC58	NGH38	394/98	MenA	MenC
287-His	32000	16	4096	4096	512	8000	16000
287(B)-His	256	-	-	-	-	16	-
287(C)-His	256	-	32	512	32	2048	>2048
287(B-C)-His	64000	128	4096	64000	1024	64000	32000

Using the domains of strain MC58, the following results were obtained: .

	MC58	2996	BZ232	NGH38	394/98	MenA	MenC
287-His	4096	32000	16	4096	512	8000	16000
287(B)-His	128	128	-	-	-	-	128
287(C)-His	-	16	-	1024	-	512	-
287(B-C)-His	16000	64000	128	64000	512	64000	>8000

Example 14 – deletions in 287

As well as expressing individual domains, 287 was also expressed (as a C-terminal His-tagged protein) by making progressive deletions within the first domain. These

Four deletion mutants of protein 287 from strain 2996 were used (Figure 6):

5 1) ‘287-His’, consisting of amino acids 18-427 (*i.e.* leader peptide deleted);
 2) ‘Δ1 287-His’, consisting of amino acids 26-427;
 3) ‘Δ2 287-His’, consisting of amino acids 70-427;
 4) ‘Δ3 287-His’, consisting of amino acids 107-427; and
 5) ‘Δ4 287-His’, consisting of amino acids 140-427 (=287-bc).
 10 The ‘Δ4’ protein was also made for strain MC58 (‘Δ4 287MC58-His’; aa 203-488).

The constructs were made in the same way as 287a/b/c, as described above.

All six constructs could be expressed and protein could be purified. Expression of 287-His was, however, quite poor.

Expression was also high when the C-terminal His-tags were omitted.

15 Immunological data (serum bactericidal assay) were also obtained using the deletion mutants, against the homologous (2996) and heterologous MenB strains, as well as MenA (F6124 strain) and MenC (BZ133 strain):

	2996	BZ232	MC58	NGH38	394/98	MenA	MenC
287-his	32000	16	4096	4096	512	8000	16000
Δ1 287-His	16000	128	4096	4096	1024	8000	16000
Δ2 287-His	16000	128	4096	>2048	512	16000	>8000
Δ3 287-His	16000	128	4096	>2048	512	16000	>8000
Δ4 287-His	64000	128	4096	64000	1024	64000	32000

The same high activity for the Δ4 deletion was seen using the sequence from strain MC58.

As well as showing superior expression characteristics, therefore, the mutants are immunologically equivalent or superior.

Example 15 – poly-glycine deletions

The ‘Δ1 287-His’ construct of the previous example differs from 287-His and from 5 ‘287^{untagged}’, only by a short N-terminal deletion (GGGGGG). Using an expression vector which replaces the deleted serine with a codon present in the *Nhe* cloning site, however, this amounts to a deletion only of (Gly)₆. Thus, the deletion of this (Gly)₆ sequence has been shown to have a dramatic effect on protein expression.

10 The protein lacking the N-terminal amino acids up to GGGGGG is called ‘ΔG 287’. In strain MC58, its sequence (leader peptide underlined) is:

→ ΔG287

1	MFKRSVIAMA CIFALSACGG GGGGSPDVKS ADTLSKPAAP VVSEKETEAK
51	EDAPQAGSQG QGAPSAQGSQ DMAAVSEENT GNNGGAVTADN PKNEDEVAQN
101	DMPQNAAGTD SSTOPNHTPDP NMLAGNMENQ ATDAGESSQP ANQPDMANAA
151	DGMQGDDPSA GGQNAGNTAA QGANQAGNMQ AAGSSDPIPA SNPAPANGGS
201	NFGRVVDLNG VLIDGPSQNI TLTHCKGDSC SGNNFLDEEV QLKSEFEKLS
251	DADKISNYKK DGKNDKFVGL VADSVQMKGI NQYIIFYKPK PTSFARFRRS
301	ARSRRSLPAE MPLIPVNQAD TLIVDGEAVS LTGHSGNIFA PEGNYRYLTY
351	GAEKLPGGSY ALRVQGEPAK GEMLAGAAVY NGEVLHFHTE NGRPYPTRGR
401	FAAKVDFGSK SVDGIIDSGD DLHMGTQKFK AAIIDGNGFKG TWTENGSGDV
451	SGKFYGPAGE EVAGKYSYRP TDAEKGGFGV FAGKKEQD*

ΔG287, with or without His-tag (‘ΔG287-His’ and ‘ΔG287K’, respectively), are expressed at very good levels in comparison with the ‘287-His’ or ‘287^{untagged}’.

25 On the basis of gene variability data, variants of ΔG287-His were expressed in *E.coli* from a number of MenB strains, in particular from strains 2996, MC58, 1000, and BZ232. The results were also good.

30 It was hypothesised that poly-Gly deletion might be a general strategy to improve expression. Other MenB lipoproteins containing similar (Gly)_n motifs (near the N-terminus, downstream of a cysteine) were therefore identified, namely Tbp2 (NMB0460), 741 (NMB 1870) and 983 (NMB1969):

→ ΔGTbp2

1	MNNPLVNQAA MVLPVFLLSA CLGGGGGSFDL DSVDTTEAPRP APKYQDVFSE
51	KPQAQKDQGG YGFAMRLKRR NWYPQAKEDE VKLDESDWEA TGLPDEPKEL
101	PKRQKSVIEK VETDSDNNIY SSPYLKPSNH QNGNTGNGIN QPKNQAKDYE
151	NPKYVYSGWF YKHAKREFNL KVEPKSAKNG DDGYIIFYHGK EPSRQLPASG
201	KITYKGVWHF ATDTKKGQKF REIIQPSKSQ GDRYSGFSGD DGEELYSNKKN
251	STLTDGQEGY GFTSNLEVDF HNKKLTGKLI RNNANTDNNQ ATTTQYYSL
301	AQVTGNRFNG KATATDKPQQ NSETKEHPFV SDSSSLSGGF FGPQGEELGF
351	RFLSDDQKVA VVGSAKTKDK PANGNTAAAS GGTDAASNG AAGTSSENGK
401	LTTVLDAVEL KLGDKEVQKL DNFNSNAAQLV VDGIMIPLL P EASESGNNQA
451	NQGTMNGGTAF TRKFDHTPES DKKDAQAGTQ TNGAQTSNT AGDTNGKTKT

501 YEVEVCCSNL NYLKYGMLTR KNSKSAMQAG ESSSQADAKT EQVEQSMFLQ
 551 GERTDEKEIP SEQNIVYRGs WYGYIANDKS TSWSGNASNA TSGNRAEFTV
 601 NFADKKITGT LTADNRQEAT FTIDGNIKDN GFEGTAKTAE SGFDLDSNT
 651 TRTPKAYITD AKVQGGFYGP KAEELGGWFA YPGDKQTKNA TNASGNSSAT
 5 701 VVFGAKRQQP VR*

741 \rightarrow $\Delta G741$
 1 VNRTAFCCCLS LTALILITAC SSGGGGVAAD IGAGLADALT APLDHDKDGL
 51 QSLTLDQSVR KNEKLKLAQ GAEKTYGNGD SLNTGKLKND KVSRFDFIRQ
 10 101 IEVDGQLITL ESGEFQVYKQ SHSALTAFQT EQIQDSEHSG KMVAKRQFRI
 151 GDIAGEHTSF DKLPEGGRAT YRGTAFGSDD AGGKLTYTID PAAKQGNNGKI
 201 EHLKSPELNV DLAAADIKPD GKRHAVISGS VLYNQAEKGS YSLGIFGGKA
 251 QEVAGSAEVK TVNGIRHIGL AAKQ*

15 983 \rightarrow $\Delta G983$
 1 MRTTPTFPTK TFKPTAMALA VATTLSACLG GGGGGTSAPD FNAGGTGIGS
 51 NSRATTAKSA AVSYAGIKNE MCKDRSMLCA GRDDVAVTDR DAKINAPPN
 101 LHTGDFPNPN DAYKNLNLK PAIEAGYTGR GVEVGIVDTG ESVGSISFPE
 151 LYGRKEHGYN ENYKNYTAYM RKEAPEDGGG KDIIEASFDD AVIETEAKPT
 201 DIRHVKEIGH IDLVSHIIGG RSVDGRPAGG IAPDATLHIM NTNDETKNEM
 251 MVAAIRNAWV KLGGERGVRIV NNSFGTTSRA GTADLFQIAN SEEQYRQALL
 301 DYSGCDKTD E GIRLMQQSDY GNLSYHIRNK NMLFIFSTGN DAQAOQPNTYA
 351 LLPFYEKDAQ KGIITVAGVD RSGEKFKREM YGEPGTEPLE YGSNHCGITA
 401 MWCLASAPYEA SVRFTRTNPI QIAGTSFSAP IVTGTAAALLL QKYPWMSMDN
 451 LRTTILLTTAQ DIGAVGVDSK FGWGLLDAGK AMNGPASFPF GDFTADTKGT
 501 SDIAYSFRND ISGTGGLIKK GGSQQLQHGN NTYTGKTIIE GGSLVLYGN
 551 KSDMRVETKG ALIYNGAASG GSLNSDGVY LADTDQSGAN ETVHIKGSLQ
 601 LDGKGTLYTR LGKLLKVDGK AIIGGKLYMS ARGKGAGYLN STGRRVPFLS
 651 AAKIGQDYSF FTNIETDGGI LASLDSVEKT AGSEGDTLSY YVRRGNAART
 701 ASAAAHSAPA GLKHAVEQGG SNLENLVEL DASESSATPE TVETAAADRT
 751 DMPGIRPYGA TFRAAAQVQH ANAADGVRIF NSLAATVYAD STAHAADM**Q**
 801 RRLKAVSDGL DHNGTGLRVI AQTQQDGGTW EQGGVEGKMR GSTQTVGIAA
 851 KTGENTTAAA TLGMGRSTWS ENSANAKTDS ISLFAGIRHD AGDIGYLKGL
 901 FSYGRYKNSI SRSTGADEHA EGSVNGTLMQ LGALGGVNVP FAATGDLTVE
 951 GGLRYDPLLKQ DAFAEKGSAL GWGNSLTERG TLVLAGLKL SQPLSDKAVL
 1001 FATAFVERDL NGRDYTVTGG FTGATAATGK TGARNMPHTR LVAGLGADVE
 1051 FGNGWNLAR YSYAGSKQYG NHSGRGRVGVGY RF*

Tbp2 and 741 genes were from strain MC58; 983 and 287 genes were from strain 2996.

40 These were cloned in pET vector and expressed in *E.coli* without the sequence coding for their leader peptides or as “ ΔG forms”, both fused to a C-terminal His-tag. In each case, the same effect was seen – expression was good in the clones carrying the deletion of the poly-glycine stretch, and poor or absent if the glycines were present in the expressed protein:

ORF	Express.	Purification	Bact. Activity
287-His(2996)	+-	+	+
'287 ^{untagged} ,(2996)	+-	nd	nd
ΔG287-His(2996)	+	+	+
ΔG287K(2996)	+	+	+
ΔG287-His(MC58)	+	+	+
ΔG287-His(1000)	+	+	+
ΔG287-His(BZ232)	+	+	+
Tbp2-His(MC58)	+-	nd	nd
ΔGTbp2-His(MC58)	+	+	
741-His(MC58)	+-	nd	nd
ΔG741-His(MC58)	+	+	
983-His (2996)			
ΔG983-His (2996)	+	+	

SDS-PAGE of the proteins is shown in Figure 13.

ΔG287 and hybrids

ΔG287 proteins were made and purified for strains MC58, 1000 and BZ232. Each of these

5 gave high ELISA titres and also serum bactericidal titres of >8192. ΔG287K, expressed from pET-24b, gave excellent titres in ELISA and the serum bactericidal assay.

ΔG287-ORF46.1K may also be expressed in pET-24b.

ΔG287 was also fused directly in-frame upstream of 919, 953, 961 (sequences shown below) and ORF46.1:

ΔG287-919

10	1	ATGGCTAGCC CCGATGTAA ATCGGCGGAC ACGCTGTCAA AACCGGCCGC
	51	TCCTGTTGTT GCTGAAAAAG AGACAGAGGT AAAAGAAGAT GCGCCACAGG
	101	CAGGGTCTCA AGGACAGGGC GCGCCATCCA CACAAGGCAG CCAAGATATG
	151	GCGGCAGTTT CGGCAGAAAA TACAGGCAAT GGCAGTGCAGG CAACAAACGGA
	201	CAAACCCAAA AATGAAGACG AGGGACCGCA AAATGATAATG CCGCAAAATT
15	251	CCGCCGAATC CGCAAATCAA ACAGGGAACA ACCAACCCGC CGATTCTTCA
	301	GATTCGGCCC CCGCGTCAA CCCTGCACCT CGGAATGGGG GTAGCAATTG
	351	TGGAAAGGGTT GATTGGCTA ATGGCGTTTT GATTGATGGG CCGTCGCAAA
	401	ATATAACGTT GACCCACTGT AAAGGCGATT CTTGTAATGG TGATAATTAA
20	451	TTGGATGAAG AAGCACCGTC AAAATCAGAA TTGAAAATT TAAATGAGTC
	501	TGAACGAATT GAGAAATATA AGAAAAGATGG GAAAAGCGAT AAATTTACTA
	551	ATTTGGTTGC GACAGCAGTT CAAGCTAATG GAACTAACAA ATATGTCATC
	601	ATTTATAAAAG ACAAGTCGC TTCATCTTCA TCTGCGCGAT TCAGGCGTTC
	651	TGCACGGTCG AGGAGGTCGC TTCCCTGCCGA GATGCCGCTA ATCCCCGTCA
	701	ATCAGGGCGGA TACGCTGATT GTCGATGGGG AAGCGGTCAAG CCTGACGGGG
25	751	CATTCCGGCA ATATCTTCGC GCGCGAAGGG AATTACCGGGT ATCTGACTTA
	801	CGGGGCGGAA AAATTGCCG GCGGATCGTA TGCCCTCCCGT GTGCAAGGCG
	851	AACCGGCAAA AGGCGAAATG CTTGCTGGCA CGGCCGTGTA CAACGGCGAA
	901	GTGCTGCATT TTCATACGGA AAACGGCCGT CGGTACCCGAA CTAGAGGCAG
	951	GTGGCCGCA AAAGTCGATT TCGGCAGCAA ATCTGTGGAC GGCATTATCG
30	1001	ACAGCGGCAGA TGATTTCGAT ATGGGTACGC AAAATTCAA AGCCGCCATC

5	1051	GATGGAAACG	GCTTTAAGGG	GAECTGGACG	GAAAATGGCG	GCAGGGATGT
	1101	TTCCGGAAGG	TTTTACGCC	CGGCCGGCGA	GGAAGTGGCG	GGAAAATACA
	1151	GCTATGCC	CACAGATGCG	GAAAAGGGCG	GATTCCGGCT	TTTGCCTGC
	1201	AAAAAAGAGC	AGGATGGATC	CGGAGGAGGA	GGATGCCAAA	GCAAGAGCAT
	1251	CCAAACCTTT	CCGCAACCCG	ACACATCCGT	CATCAACGGC	CCGGACCGGC
	1301	CGGTCGGCAT	CCCCGACCCC	GCCGGAACGA	CGGTCGGCGG	CGGCAGGGCC
	1351	GTCTATACCG	TTGTACCGCA	CCTGTCCCTG	CCCCACTGGG	CGGCAGGAGA
	1401	TTTCGCCAAA	AGCCTGCAAT	CCTTCCGCCT	CGGCTGCGCC	AATTTGAAAA
10	1451	ACCGCCAAGG	CTGGCAGGAT	GTGTGCGCCC	AAGCCTTCA	AACCCCCGTC
	1501	CATTCCCTTC	AGGCAAACAA	GTTCAGGAA	CGCTATTTCA	CGCCGTGGCA
	1551	GGTTGCAGGC	AACGGAAGCC	TTGCCGGTAC	GGTTACCGGC	TATTACGAGC
	1601	CGGTGCTGAA	GGGCGACGAC	AGGCGGACGG	CACAAGCCCG	CTTCCCGATT
	1651	TACGGTATTTC	CCGACGATT	TATCTCCGTC	CCCTGCCTG	CCGGTTTGC
15	1701	GAGCGGAAAAA	GCCCTTGCC	GCATCAGGCA	GACGGGAAAAA	AACAGCGGCA
	1751	CAATCGACAA	TACCGGGCG	ACACATACCG	CCGACCTCTC	CCGATTCCCC
	1801	ATACCCGCGC	GCACAAACGGC	AATCAAAGGC	AGGTTTGAAG	GAAGCCGCTT
	1851	CCTCCCTTAC	CACACCGCGA	ACCAAATCAA	CGGCGGCGCG	CTTGACGGCA
	1901	AAGCCCCGAT	ACTCGGTTAC	GCCGAAGACC	CCGTCGAAC	TTTTTTTATG
20	1951	CACATCCAAG	GCTCGGGCCG	TCTGAAAACC	CCGTCGGCA	AATACATCCG
	2001	CATCGGCTAT	GCCGACAAAAA	ACGAACATCC	CTACGTTTCC	ATCGGACGCT
	2051	ATATGGCGGA	CAAAGGCTAC	CTCAAGCTCG	GGCAGACCTC	GATGCAGGGC
	2101	ATCAAAGCCT	ATATGCGCA	AAATCCGCAA	CGCCTCGCCG	AAGTTTTGGG
	2151	TCAAAACCCCC	AGCTATATCT	TTTCCCGCGA	GCTTGCCGGA	ACCACCAATG
25	2201	ACGGTCCCGT	CGGCGCACTG	GGCACGCGT	TGATGGGGGA	ATATGCCGGC
	2251	GCAGTCGACC	GGCACTACAT	TACCTTGGGC	GGCGCCCTTAT	TTGTCGCCAC
	2301	CGCCCATCCG	GTTACCCGCA	AAAGCCTCAA	CCGCTCTGATT	ATGGCGCAGG
	2351	ATACCGGCAG	CGCGATTAAA	GGCGCGGTGC	GGTGGGATTA	TTTTGGGGGA
	2401	TACGGCGACG	AAGCCGGCGA	ACTTGCCGGC	AAACAGAAAAA	CCACGGGTTA
	2451	CGTCTGGCAG	CTCCTACCA	ACGGTATGAA	GCCCGAATAC	CGCCCGTAAC
30	2501	TCGAG				
	1	MASPDVKSAD	TLSKPAAPVV	AEKETEVKED	APQAGSQGQG	APSTQGSQDM
	51	AAVSAENTGN	GGAATTDKPK	NEDEGPQNDM	PQNSAESANQ	TGNNQPADSS
35	101	DSAPASNPAP	ANGGSNFGRV	DLANGVLIDG	PSQNITLTHC	KGDSCMNDNL
	151	LDEEAPSKSE	FENLNESERI	EKYKDKGKSD	KFTNLVATAV	QANGTNKYVI
	201	IYKDKSASSS	SARFRRSARS	RRSLPAEMPL	IPVNQADTLI	VDGEAVSLTG
	251	HSGNIFAPEG	NYRYLTYGAE	KLPGGSYALR	VQGEPAKgem	LAGTAVYNGE
	301	VLIHFHTENGR	PYPTTRGRFAA	KVDFGSKSVD	GIIDSGDDLH	MGTQKFKAII
40	351	DGNGFKGTWT	ENGGGDVSGR	FYGPAGEEVA	GKYSYRPTDA	EKGFFGVFAG
	401	KKEQDGSGGG	GCQSKSIQTF	PQPDTSVING	PDRPVGIPDP	AGTTVGGGA
	451	VYTVPVPHLSL	PHWAAQDFAK	SLQSFRLGCA	NLKNRQGWQD	VCAQAFQTPV
	501	HSFQAKQFFE	RYFTPWQVAG	NGSLAGTVG	YYEPVULKGD	RRTAQARFPI
	551	YGPDDFISV	PLPAGLRSGK	ALVRIQRTGK	NSGTIDNTGG	THTADLSRFP
45	601	ITARTTAIKG	RFEGRFLP	HTRNQINGGA	LDGKAPILGY	AEDPVELFFM
	651	HIQGSGLKLT	PSGKYIRIGY	ADKNEHPYVS	IGRYMADKGY	LKLGQTSMQG
	701	IKAYMRQNPQ	RLAEVLGQNP	SYIFRELAG	SSNDGPVGAL	GTPLMGEYAG
	751	AVDRHYITLG	APLFVATAHP	VTRKALNRLI	MAQDTGSAIK	GAVRVDYFWG
	801	YGDEAGELAG	KQKTTGYVWQ	LLPNGMKPEY	RP*	
50						
		ΔG287-953				
	1	ATGGCTAGCC	CCGATGTTAA	ATCGCGGGAC	ACGCTGTCAA	AACCGGCCGC
	51	TCCGTGTTGTT	GCTAAAAAG	AGACAGAGGT	AAAAGAAGAT	GCGCCACAGG
55	101	CAGGTTCTCA	AGGACAGGGC	GCGCCATCCA	CACAAGGCAG	CCAAGATATG
	151	GCGGCAGTTT	CGGCAGAAAAA	TACAGGCAAT	GGCGGTGCGG	CAACAAACGGA
	201	CAAACCCAAA	AATGAAGACG	AGGGACCGCA	AAATGATATG	CCGCAAAATT
	251	CCGCCGAATC	CGCAAATCAA	ACAGGAAACA	ACCAACCCCGC	CGATTCTTC
	301	GATTCCGCC	CCGCGTCAA	CCCTGCACCT	CGAATGGCG	GTAGCAATT
60	351	TGGAAGGGTT	GATTTGGCTA	ATGGCGTTT	GATTGATGGG	CCGTCGCAAA
	401	ATATAACGTT	GACCCACTGT	AAAGGCGATT	CTTGTAAATGG	TGATAATTAA
	451	TTGGATGAAAG	AAGCACCGTC	AAAATCAGAA	TTTGAAGATT	AAATGAGTC
	501	TGAACGAATT	GAGAAATATA	AGAAAGATGG	GAAAAGCGAT	AAATTTACTA
	551	ATTGGTTGC	GACAGCAGTT	CAAGCTAATG	GAACTAACAA	ATATGTCATC
65	601	ATTTATAAAG	ACAAGTCGC	TTCATCTTCA	TCTGCGCGAT	TCAGGCGTTC
	651	TGCACGGTCG	AGGAGGTGCG	TTCCCTGCCGA	GATGCCGCTA	ATCCCCGTC
	701	ATCAGGCGGA	TACGCTGATT	GTGCGATGGGG	AAGCGGTCAG	CCTGACGGGG
	751	CATTCCGGCA	ATATCTTCGC	GCCCGAAGGG	AATTACCGGT	ATCTGACTTA

801	CGGGGCGGAA	AAATTGCCCG	GCGGATCGTA	TGCCCTCCGT	GTGCAAGGCG	
851	AACCGGCAAA	AGGCAGAAATG	CTTGCCTGGCA	CGGCCGTGTA	CAACGGCGAA	
901	GTGCTGCAATT	TTCATACCGA	AAACGGCCGT	CCGTACCCGA	CTAGAGGCAG	
951	GTGGCCGCA	AAAGTCGATT	TCGGCAGCAA	ATCTGTGGAC	GGCATTATCG	
5	1001	ACAGCGGCGA	TGATTTGCAT	ATGGGTACGC	AAAATTCAA	AGCCGCCATC
	1051	GATGGAAACG	GCTTTAAGGG	GACTTGGACG	GAAAATGGCG	CGGGGGATGT
	1101	TTCCGGAAGG	TTTACGGCC	CGGCCGGCGA	GGAAGTGGCG	GGAAAATACA
	1151	GCTATCGCCC	GACAGATGCG	GAAAAGGGCG	GATTCCGGGT	TTTGCCGGC
10	1201	AAAAAAAGAGC	AGGATGGATC	CGGAGGGAGGA	GGAGCCACCT	ACAAAGTGGA
	1251	CGAATATCAC	GCCAACGCC	GTTTCGCCAT	CGACCATTTC	AACACCAGCA
	1301	CCAACGTCGG	CGGTTTTAC	GGTCTGACCG	GTTCCGTCGA	GTTCGACCAA
	1351	GCAAAACGCG	ACGGTAAAT	CGACATCACC	ATCCCCGTTG	CCAACCTGCA
	1401	AAGCGGTTCG	CAACACTTA	CCGACCACCT	GAAATCAGCC	GACATCTTCG
15	1451	ATGCCGCCCA	ATATCCGAC	ATCCGCTTTG	TTTCCACCAA	ATTCAACTTC
	1501	AACGGCAAAA	AACTGGTTTC	CGTTGACCGC	AACTGACCA	TGCACGGCAA
	1551	AACCGCCCCC	GTCAAACCTA	AAGCCGAAAA	ATTCAACTGC	TACCAAAGCC
	1601	CGATGGCGAA	AACCGAAGTT	TGCGGCGGCG	ACTTCAGCAC	CACCATCGAC
	1651	CGCACCAAAAT	GGGGCGTGGA	CTACCTCGTT	AACGTTGGTA	TGACCAAAAG
	1701	CGTCCGCATC	GACATCCAAA	TCGAGGCAGC	CAAACAATAA	CTCGAG
20	1	MASPDVKSAD	TLSKPAPVV	AEKETEVKED	APQAGSQGQG	APSTQGSQDM
	51	AAVSAENTGN	GGAATTDKPK	NEDEGPQNDM	PQNSAESANQ	TGNNQPADSS
	101	DSAPASNPA	ANGGSNFGRV	DLANGVLIDG	PSQNITLTHC	KGDSCNGDNL
25	151	LDEEAPSKSE	FENLNESERI	EKYKKDGKSD	KFTNLVATAV	QANGTNKYVI
	201	IYKDKSASSS	SARFRRSARS	RRSLPAEMPL	IPVNQADTLI	VDGEAVSLTG
	251	HSGNIFIPEG	NYRILTYGARS	KLPGGSYALR	VQGEPAKGEML	LAGTAVYNGE
	301	VLHFHTENGR	PYPTTRGRFAA	KVDFGSKSVD	GIIDSGDDLH	MGTQKFKAII
	351	DGMNGFKGTWT	ENGGGDVSGR	FYGPAGEEEVA	GKYSYRPTDA	EKGFFGVFAG
30	401	KKEQDGSGGG	GATYKVDEYH	ANARFAIDHF	NTSTNVGGFY	GLTGSVEFDQ
	451	AKRDGKIDIT	IPVANLQSGS	QHFTDHLKSA	DIFDAAQYPD	IRFVSTKFNF
	501	NGKKLVSVDG	NLTMHGKTAP	VKLKAEKFNC	YQSPMAKTEV	CGGDFSTTID
	551	RTKWDYLV	NVGMTKSVRI	DIQIEAAKQ*		
35	<u>AG287-961</u>					
	1	ATGGCTAGCC	CCGATTTAA	ATCGGGGAC	ACGCTGTCAA	AACCGGCCGC
	51	TCCCTGTTGTT	GCTAAAAAG	AGACAGAGGT	AAAAGAAGAT	GCGCACAGG
	101	CAGGTTCTCA	AGGACAGGGC	GCGCCATCCA	CACAAGGGCAG	CCAAGATATG
40	151	GCGGAGTTT	CGGCAGAAAA	TACAGGCAAT	GGCGGTGCGG	CAACAACGGA
	201	CAAACCCAAA	AATGAAGACG	AGGGACCGCA	AAATGATATG	CCGAAAATT
	251	CCGCCGAATC	CGCAAATCAA	ACAGGGAAACA	ACCAACCCCGC	CGAFTCTTC
	301	GATCCGCC	CCGCGTCAAA	CCCTGACCT	GCGAATGGCG	GTAGCAATT
	351	TGGAAGGGT	GATTTGGCTA	ATGGCGTTT	GATTGATGGG	CCGTCGAAA
45	401	ATATAACGTT	GACCCACTGT	AAAGGGCATT	CTTGTAAATGG	TGATAATT
	451	TTGGATGAAG	AAGCACCGTC	AAAATCAGAA	TTTGAAAATT	TAATGAGTC
	501	TGAACGAATT	GAGAAATATG	AGAAAGATGG	AAAAGCGAT	AAATTACTA
	551	ATTGGTTGC	GACAGCAGTT	CAAGCTAATG	GAACTAACAA	ATATGTCATC
	601	ATTTATAAAG	ACAAGTCCGC	TTCATCTTC	TCTGCGCGAT	TCAGGCGTTC
50	651	TGCACGGTCG	AGGAGGTCGC	TTCCCTGCCGA	GATGCCGCTA	ATCCCCGTCA
	701	ATCAGGCGGA	TACGCTGATT	GTCGATGGGG	AAGCGGTAG	CCTGACGGGG
	751	CATTCCGGCA	ATATCTTCG	GCCCCGAGGG	AATTACCGGT	ATCTGACTTA
	801	CGGGCGGGAA	AAATTGCGCG	GCGGATCGTA	TGCCCTCCGT	GTGCAAGGCG
	851	AACCGGCAA	AGGCAATG	CTTGCCTGGCA	CGGCCGTGTA	CAACGGCGAA
55	901	GTGCTGCAATT	TTCATACCGA	AAACGGCCGT	CCGTACCCGA	CTAGAGGCAG
	951	GTGGCCGCA	AAAGTCGATT	TCGGCAGCAA	ATCTGTGGAC	GGCATTATCG
	1001	ACAGCGGCGA	TGATTTGCAT	ATGGGTACGC	AAAATTCAA	AGCCGCCATC
	1051	GATGGAAACG	GCTTTAAGGG	GACTTGGACG	GAAAATGGCG	CGGGGGATGT
	1101	TTCCGGAAGG	TTTACGGCC	CGGCCGGCGA	GGAAGTGGCG	GGAAAATACA
60	1151	GCTATCGCCC	GACAGATGCG	GAAAAGGGCG	GATTCCGGGT	TTTGCCGGC
	1201	AAAAAAAGAGC	AGGATGGATC	CGGAGGGAGGA	GGAGCCACAA	ACGACGACGA
	1251	TGTTAAAAAA	GCTGCCACTG	TGGCCATTGC	TGCTGCTTAC	AACAATGGCC
	1301	AAGAAATCAA	CGGTTTCAA	GCTGGAGAGA	CCATCTACGA	CATTGATGAA
	1351	GACGGCACAA	TTACCAAAA	AGACGCAACT	GCAGCCGATG	TTGAAGCCGA
65	1401	CGACTTTAAA	GGTCTGGGT	TGAAAAAAAGT	CGTGACTAAC	CTGACCAAAA
	1451	CCGTCAATGA	AAACAAACAA	AACGTCGATG	CCAAAGTAA	AGCTGCAGAA
	1501	TCTGAAATAG	AAAAGTTAAC	AACCAAGTTA	GCAGACACTG	ATGCCGCTTT
	1551	AGCAGATACT	GATGCCGCTC	TGGATGCAAC	CACCAACGCC	TTGAATAAT

1601 TGGGAGAAAA TATAACGACA TTTGCTGAAG AGACTAAGAC AAATATCGTA
 1651 AAAATTGATG AAAAATTAGA AGCCGTGGCT GATACCGTCG ACAAGCATGC
 1701 CGAACGATTC AACGATATCG CCGATTCTATT GGATGAAAC AACACTAAAGG
 1751 CAGACGAAGC CGTCAAAACC GCCAATGAAG CAAACAGAC GGCGAAGAA
 5 1801 ACCAAACAAA ACGTCGATGC CAAAGTAAAA GCTGCAGAAA CTGCAGCAGG
 1851 CAAAGCCGAA GCTGCCGCTG GCACAGCTAA TACTGCAGCC GACAAGGCCG
 1901 AAGCTGTCGC TGCAAAAGTT ACCGACATCA AAGCTGATAT CGCTACGAAC
 1951 AAAGATAATA TTGCTAAAAA AGCAAAACAGT GCGCACGTGT ACACCCAGACA
 10 2001 AGAGTCTGAC AGCAAATTG TCAGAATTGA TGGTCTGAAC GCTACTACCG
 2051 AAAAATTGGA CACACGCTTG GCTCTGCTG AAAAATCCAT TGCCGATCAC
 2101 GATACTCGCC TGAACGGTT GGATAAAACA GTGTCAGACC TGCGCAAAGA
 2151 AACCCGCCAA GGCCTTGCAAG AACAGCCGC GCTCTCCGGT CTGTTCCAAC
 2201 CTTACAAACGT GGGTCGGTTC AATGTAACGG CTGCAGTCGG CGGCTACAAA
 2251 TCCGAATCGG CAGTCGCCAT CGGTACCGGC TTCCGCTTTA CCGAAAACCTT
 15 2301 TGCCGCCAAA GCAGGCGTGG CAGTCGGCAC TCTCGTCCGGT TCTTCCGCAG
 2351 CCTACCATGT CGCGTCAAT TACGAGTGGT AACTCGAG

1 MASPDVKSAD TLSKPAAAPVV AEKETEVKED APQAGSQGQG APSTQGSQDM
 20 51 AAVSAENTGN GGAATTDKPK NEDEGPQNDM PQNSAESANQ TGNNQPADSS
 101 DSAPASNPA PANGGSNFGRV DLANGVLIIDG PSQNIITLTHC KGDSCNGDNL
 151 LDEEAPSKSE FENLNESERI EKYKKDGKSD KFTNLVATAV QANGTNKYVI
 201 IYKDKSASSS SARFRRSARS RRSLPAEMPL IPVNQADTLI VDGEAVSLTG
 251 HSGNIFAPEG NYRYLTYGAE KLPGGSYALR VQGEPAKGEM LAGTAVYNGE
 301 VLHFHTENGR PYPTGRFRAA KVDFGSKSVD GIIDSGDDIH MGTQKFKAII
 351 DGMGFKGTWT ENGGGDVSGR FYGPAGEEVA GKYSYRPTDA EKGGFGVFAG
 401 KKEQDGSGGG GATNDDDVKK AATVAIAAAAY NNGQEINGFK AGETIYDIDE
 451 DGTITKKDADT AADVEADDK GLGLKKVVTN LTKTVMENKQ NVDAKVAAE
 501 SEIEKLTTKL ADTDAALADT DAALDATTNA LNKLGENITT FAEETKTNIV
 551 KIDEKLEAVA DTVDKHAEEF NDIADSLDET NTKADEAVKT ANEAKQTAEE
 30 601 TKQNVDAVK AAETAAGKAE AAAGTANTAA DKAEEAVAAKV TDIKADIATN
 651 KDNIAKKKANS ADVYTREESD SKFVRIDGLN ATTEKLDTRL ASAEKSIADH
 701 DTRLNGLDKT VSDLRKETRQ GLAEQAALSG LFQPYNVGRF NVTAAVGGYK
 751 SESAVAIGTG FRFTENFAAK AGVAVGTSSG SSAAYHVGVN YEW*

	ELISA	Bactericidal
ΔG287-953-His	3834	65536
ΔG287-961-His	108627	65536

35 The bactericidal efficacy (homologous strain) of antibodies raised against the hybrid proteins was compared with antibodies raised against simple mixtures of the component antigens (using 287-GST) for 919 and ORF46.1:

	Mixture with 287	Hybrid with ΔG287
919	32000	128000
ORF46.1	128	16000

Data for bactericidal activity against heterologous MenB strains and against serotypes A and C were also obtained:

Strain	919		ORF46.1	
	<i>Mixture</i>	<i>Hybrid</i>	<i>Mixture</i>	<i>Hybrid</i>
NGH38	1024	32000	-	16384
MC58	512	8192	-	512
BZ232	512	512	-	-
MenA (F6124)	512	32000	-	8192
MenC (C11)	>2048	>2048	-	-
MenC (BZ133)	>4096	64000	-	8192

The hybrid proteins with ΔG287 at the N-terminus are therefore immunologically superior to simple mixtures, with ΔG287-ORF46.1 being particularly effective, even against heterologous strains. ΔG287-ORF46.1K may be expressed in pET-24b.

The same hybrid proteins were made using New Zealand strain 394/98 rather than 2996:

5 ΔG287NZ-919

1	ATGGCTAGCC	CCGATGTCAA	GTCGGCGGAC	ACGCTGTCAA	AACCTGCCGC
51	CCCTGTTGTT	TCTGAAAAAG	AGACAGAGGC	AAAGGAAGAT	GCGCCACAGG
101	CAGGTTCTCA	AGGACAGGGC	GCGCCATCCG	CACAAGGC GG	TCAAGATATG
151	GCGGCGGTTT	CGGAAGAAAA	TACAGGCAAT	GGCGGTGCGG	CAGCAACGGA
201	CAAACCCAAA	AATGAAGACG	AGGGGGCGCA	AAATGATATG	CCGCAAAATG
251	CCGCGATAC	AGATAGTTTG	ACACCGAATC	ACACCCCGGC	TTCGAATATG
301	CCGGCCGGAA	ATATGGAAA	CCAAGCACC	GATGCCGGGG	AATCGGAGCA
351	GCCGGCAAC	CAACCGGATA	TGGCAAATAC	GGCGGACGGA	ATGCAGGGTG
401	ACGATCCGTC	GGCAGGGGG	GAAAATGCCG	GCAATACGGC	TGCCCCAAGGT
451	ACAAATCAAG	CCGAAAACAA	TCAAACCGCC	GGTCTCAAA	ATCCTGCCCTC
501	TTCAACCAAT	CCTAGGCCA	CGAATAGCGG	TGGTGATTTT	GGAAGGACGA
551	ACGTGGGCAA	TTCTGTTG	ATTGACGGGC	CGTCGCAAA	TATAACGTTG
601	ACCCACTGTA	AAGGCGATT	TTGTAGTGGC	AATAATTCT	TGGATGAAAGA
651	AGTACAGCTA	AAATCAGAA	TTGAAAATT	AAAGTATGCA	GACAAAATAA
701	GTAATTACAA	GAAAGATGGG	AAGAATGACG	GGAAGAATGA	TAAATTGTC
751	GGTTTGGTTG	CCGATAGTGT	GCAGATGAAG	GGAATCAATC	AATATATTAT
801	CTTTTATAAA	CCTAAACCCA	CTTCATTG	GCGATTTAGG	CGTTCTGCAC
851	GGTCGAGGCC	GTCGCTTCG	GGCGAGATGC	CGCTGATTCC	CGTCAATCAG
901	GCGGATACGC	TGATTGTCGA	TGGGGAAGCG	GTCAGCCTGA	CGGGGCATT
951	CGGCAATATC	TTCGCGCCCG	AAGGGAATT	CCGGTATCTG	ACTTACGGGG
1001	CGGAAAATT	GCCCCGGGA	TCGTATGCC	TCCGTGTTCA	AGGCGAACCT
1051	TCAAAAGGCC	AAATGCTCG	GGGCACGGCA	GTGTACAACG	GCGAAGTGCT
1101	GCATTTCAT	ACGGAAAACG	GCCGTCCGTC	CCCGTCCAGA	GGCAGGTTTG
1151	CCGCAAAAGT	CGATTCGGC	AGCAAATCTG	TGGACGGCAT	TATCGACAGC
1201	GGCGATGGTT	TGCATATGGG	TACGAAAAA	TTCAAAGCCG	CCATCGATGG
1251	AAACGGCTTT	AAGGGGACT	GGACGGAAA	TGGCGGGCGG	GATGTTCCG
1301	GAAAGTTTA	CGGCCCGGCC	GGCGAGGAAG	TGGCGGGAAA	ATACAGCTAT
1351	CGGCCAACAG	ATGCGGAAA	GGGGGGATT	GGCGTGTGTTG	CCGGCAAAA
1401	AGAGCAGGAT	GGATCCGGAG	GAGGAGGATG	CCAAAGCAAG	AGCATCCAA
1451	CCTTTCCGCA	ACCCGACACA	TCCGTCA	ACGGCCCCGG	CCGGCCGGTC
1501	GGCATCCCCG	ACCCCGCCG	AACGACGGTC	GGCGGGCGG	GGGCCGTCTA
1551	TACCGTTGTA	CCGCACCTGT	CCCTGCCCCA	CTGGGGCGG	CAGGATTTCG
1601	CCAAAAGCCT	GCAATCCTTC	CGCCTCGGCT	GCGCCAATT	GAAAACCGC
1651	CAAGGCTGGC	AGGATGTGT	CGCCCAAGCC	TTTCAAACCC	CCGTCCATT
1701	CTTCAGGCA	AAACAGTTT	TTGAACGCTA	TTTCACGCCG	TGGCAGGTTG
1751	CAGGCAACGG	AAGCCTTGCC	GGTACGGTTA	CCGGCTATT	CGAGCCGGT
1801	CTGAAGGGCG	ACGACAGGCG	GACGGCACAA	GCCCGCTTCC	CGATTTACGG
1851	TATTCCCGAC	GATTTATCT	CCGTC	CCGCTGCGG	TTGCGGAGCG
1901	GAAAGCCCT	TGTCCGCATC	AGGCAGACGG	GAAAAAACAG	CGGCACAATC
1951	GACAATACCG	GCGGCACACA	TACCGCCGAC	CTCTCCC	GAT TCCCCATCAC
2001	CGCGCGCACA	ACGGCAATCA	AAGGCAGGTT	TGAAGGAAGC	CGCTTCCCTCC

-40-

2051	CCTACCACAC	GCGCAACCAA	ATCAACGGCG	GCGCGCTTGA	CGGCAAAGCC	
2101	CCGATACTCG	GTTACGCCGA	AGACCCCGTC	GAACCTTTTT	TTATGCACAT	
2151	CCAAGGCTCG	GGCCGTCTGA	AAACCCCGTC	CGGCAAATAC	ATCCGCATCG	
2201	GCTATGCCGA	AAAAAACGAA	CATCCCTACG	TTTCCATCGG	ACGCTATATG	
2251	CGGGACAAAG	GCTACCTCAA	GCTCGGGCAG	ACCTCGATGC	AGGGCATTCAA	
2301	AGCCTATATG	CGGCAAAATC	CGCAACGCCT	CGCCGAAGTT	TTGGGTCAAA	
2351	ACCCCAGCTA	TATCTTTTC	CGCGAGCTTG	CGGGAAGCAG	CAATGACGGT	
2401	CCCGTCCGGCG	CACTGGGCAC	GCCGTTGATG	GGGGAATATG	CGGGCGCAGT	
2451	CGACCGGCAC	TACATTACCT	TGGGCGCGCC	CTTATTTGTC	GCCACCGCCC	
10 2501	ATCCGGTTAC	CGGCAAAGCC	CTCAACCGCC	TGATTATGGC	GCAGGATACC	
2551	GGCAGCGCGA	TTAAAGGCGC	GGTGCAGCTG	GATTATTTT	GGGGATAACGG	
2601	CGACGAAGCC	GGCGAACTTG	CGGCAAACAA	GAAAACCACG	GGTTACGTCT	
2651	GGCAGCTCCT	ACCCAACGGT	ATGAAGCCCG	AATACCGCCC	GTAAAAGCTT	
15 1	MASPDVKSAD	TLSKPAPVV	SEKETEAKED	APQAGSQGQG	APSAQGGQDM	
51	AAVSEENTGN	GGAAATDKPK	NEDEGAQNDM	PQNAADTDSL	TPNHTPASN	
101	PAGNMENQAP	DAGESEQPAN	QPDMANTADG	MQGDDPSAGG	ENAGNTAAQG	
151	TNQAEENNQTA	GSQNPASSTN	PSATNSGGDF	GRTNVGNNSV	IDGFSQNTIL	
20 201	THCKGDSCSG	NNFLDEEVQL	KSEFEKLSDA	DKISNYKKDG	KNDGKNDKFV	
251	GLVADSVQMK	GINQYIIFYK	PKPTSFARFR	RSARSRRSLP	AEMPLIPVNQ	
301	ADTLIVDGEA	VSLTGHSGNI	FAPEGNYRYL	TYGAEKLPGG	SYALRVQGEP	
351	SKGEMLLAGTA	VYNGEVLHFH	TENGRPSPSR	GRFAAKVDFG	SKSVDGIIIDS	
401	GDGLHMGCTQR	FKAAIDGNF	KGTWTENGGG	DVSGKFYGP	GEEVAGKYSY	
451	RPTDAEKGGF	GVFAGKKEQD	GSGGGCQSK	SIQTFPQPDT	SVINGPDRPV	
501	GIPDPAGTTV	GGGGAVYTVV	PHLSLPHWAA	QDFAKSLQSF	RLGCANLKRN	
551	QGWQDVCAQA	FQTPVHSFQA	KQFFERYFTP	WQVAGNGSLA	GTVTGYYEPV	
601	LKGDDRTAQ	ARFPIYGIPD	DFISVPLPAG	LRSGKALVRI	RQTGKNSGTI	
651	DNTGGTHTAD	LSRFPIATART	TAIKGRFEGS	RFLPYHTRNQ	INGGALDGKA	
30 701	PILGYAEDPV	ELFFMHIQGS	GRLKTPSGKY	IRIGYADKNE	HPYVSIGRYM	
751	ADKGYLKLGQ	TSMQGIKAYM	RQNQRLAEV	LGQNPSYIFF	RELAGSSNDG	
801	PVGALGTPLM	GEYAGAVDRH	YITLGAPLFV	ATAHPVTRKA	LNRLIMAQDT	
851	GSATKGAVRV	DYFWGYGDEA	GELAKQKTT	GYVWQLLPNG	MKPEYRP*	
35	<u>AG287NZ-953</u>					
1	ATGGCTAGCC	CCGATGTCAA	GTCGGCGGAC	ACGCTGTCAA	ACCTGCCGC	
51	CCCTGTTGTT	TCTGAAAAAG	AGACAGAGGC	AAAGGAAGAT	GCGCCACAGG	
101	CAGGGTCTCA	AGGACAGGGC	GCGCCATCCG	CACAAGGCGG	TCAAGATATG	
151	CGGGCGGTTT	CGGAAGAAAA	TACAGGCAAT	GGCGGTGCGG	CAGCAACGGA	
20 201	CAAACCCAAA	AATGAAGACG	AGGGGGCGCA	AAATGATATG	CCGAAAATG	
251	CCGCCGATAC	AGATAGTTTG	ACACCGAATC	ACACCCCGGC	TTCGAATATG	
301	CCGGCCGGAA	ATATGGAAA	CCAAGCACC	GATGCCGGGG	AATCGGAGCA	
351	GCCGGCAAAC	CAACCGGATA	TGGCAAATAC	GGCGGACCGG	ATGCAGGGTG	
401	ACGATCCGTC	GGCAGGCCGG	AAAAATGCCG	GCAATACGGC	TGCCCCAAGGT	
451	ACAAATCAAG	CCGAAAACAA	TCAAACGCC	GGTTCTCAA	ATCCTGCCTC	
501	TTCAACCAAT	CCTAGGCCA	CGAATAGCGG	TTGTGATT	GGAAGGACGA	
551	ACGTGGCAA	TTCTGTTGTG	ATTGACGGGC	CGTCGCAAA	TATAACGTTG	
601	ACCCACTGTA	AAGGCGATTC	TTGTAGTGGC	AATAATTCT	TGGATGAAGA	
651	AGTACAGCTA	AAATCAGAA	TTGAAAAATT	AAAGTATGCA	GACAAAATAA	
701	GTAATTACAA	GAAAGATGGG	AAGAATGACG	GGAAGAATGA	TAAATTGTC	
751	GGTTTGGTTG	CCGATAGTGT	GCAGATGAAG	GGAATCAATC	AATATATTAT	
801	CTTTTATAAA	CCTAAACCCA	CTTCATTGTC	GCGATTTAGG	CGTTCTGCAC	
851	GGTCGAGGCC	GTCGCTTCG	GCGCAGATCC	CGCTGATTCC	CGTCAATCAG	
901	CGGGATAACGC	TGATTGTCGA	TGGGGAAGCG	GTCAAGCTGA	CGGGGCATT	
951	CGGCAATATC	TTCGCGCCCG	AAGGGAATT	CCGGTATCTG	ACTTACGGGG	
1001	CGGAAAATT	GCCCCGGGA	TCGTATGCC	TCCGTGTTCA	AGGCAGAACCT	
1051	TCAAAAGGCG	AAATGCTCG	GGGCACGGCA	GTGTACAACG	GCGAAGTGCT	
1101	GCATTTTCAT	ACGGAAAACG	GCCGTCCGTC	CCCCTCCAGA	GGCAGGTTTG	
1151	CCGCAAAAGT	CGATTTCGGC	AGCAAATCTG	TGGACGGCAT	TATCGACAGC	
1201	GGCGATGGTT	TGCATATGGG	TACGCAAAA	TTCAAAGCCG	CCATCGATGG	
1251	AAACGGCTTT	AAGGGGACTT	GGACGGAAA	TGGCGGCCGG	GATGTTCCG	
1301	GAAAGTTITA	CGGCCCCGCC	GGCGAGGAAG	TGGCGGGAAA	ATACAGCTAT	
1351	CGCCCAACAG	ATGCGGAAA	GGCGGGATT	GGCGTGT	CCGGCAAAAA	
1401	AGAGCAGGAT	GGATCCGGAG	GAGGAGGAGC	CACCTACAAA	GTGGACGAAT	
1451	ATCACGCCAA	CGCCCGTTTC	GCCATCGACC	ATTTCAACAC	CAGCACCAAC	
1501	GTCGGCGGTT	TTTACGGTCT	GACCGGTTCC	GTCAAGTTCG	ACCAAGCAAA	
1551	ACCGCAGCGGT	AAAATCGACA	TCACCATCCC	CGTTGCCAAC	CTGCAAAGCG	

-41-

1601	GTTCGCAACA	CTTTACCGAC	CACCTGAAAT	CAGCCGACAT	CTTCGATGCC	
1651	GCCAATATC	CGGACATCCG	CTTTGTTTCC	ACCAAATTCA	ACTTCAACGG	
1701	CAAAAAACTG	GTTTCCGTTG	ACGGCAACCT	GACCATGCAC	GGCAAAACCG	
1751	CCCCCGTCAA	ACTCAAAGCC	GAAAATTCA	ACTGCTACCA	AAGCCCGATG	
5	1801	GGAAAACCG	AAGTTGCGG	CGGCGACTTC	AGCACCACCA	TCGACCGCAC
1851	CAAATGGGGC	GTGGACTACC	TCGTTAACGT	TGGTATGACC	AAAAGCGTCC	
1901	GCATCGACAT	CCAAATCGAG	GCAGCCTAAC	AATAAAAGCT	T	
10	1	MASPDVKSAD	TLSKPAAFPVV	SEKETEAKED	APQAGSQGQG	APSAQGGQDM
	51	AAVSEENTGN	GGAAATDKPK	NEDEGAQNDM	PQNAADTDLS	TPNHTPASN
	101	PAGNMENQAP	DAGESEQPAN	QPDmantADG	MQGDDPSAGG	ENAGNTAAQG
	151	TNQAENNQTA	GSQNPASTN	PSATNSCGDF	GRTNVGNNSV	IDGPSQNITL
	201	THCKGDSCSG	NNFLDEEVQL	KSEFEKLSDA	DKISNYKKDG	KNDGKNDKFV
15	251	GLVADSVQMK	GINQYIIFYK	PKPTSFARFR	RSARSRRSLP	AEMPLIPVNQ
	301	ADTLIVDGEA	VSLTGHSGNI	FAPEGNYRYL	TYGAEKLPGG	SYALRVQGEP
	351	SKGEMLAGTA	VYNGEVLHNFH	TENGRRPSR	GRFAAKVDFG	SKSVDGIIDS
	401	GDGLHMGQTQK	FKAAIDGNGF	KGTWTFENGGG	DVSGKFYGP	GEEVAGKSY
	451	RPTDAEKGGF	GVFAGKKEQD	GSGGGGATYK	VDEYHANARF	AIDHFNTSTN
20	501	VGGFYGLTGS	VEFDQAKRDG	KIDITIPVAN	LQSGSQHFTD	HLKSADIFDA
	551	AQYPDIRFVS	TKFNFNGKKL	VSDGNLTMH	GKTAPVKLKA	EKFNCYQSPM
	601	AKTEVCGGDF	STTIDRTKVG	VDYLVNVGMT	KSVRIDIQIE	AAKQ*

AG287NZ-961

25	1	ATGGCTAGCC	CCGATGTCAA	GTCGGCGGAC	ACGCTGTCAA	AACCTGCCGC
	51	CCCTGTTGTT	TCTGAAAAG	AGACAGAGGC	AAAGGAAGAT	GCGCACAGG
	101	CAGGTTCTCA	AGGACAGGGC	GCGCCATCCG	CACAAGGCGG	TCAAGATATG
	151	GCGCGGTTT	CGGAAGAAAA	TACAGGCAAT	GGCAGTGC	CAGCAACGGA
30	201	CAAACCCAAA	AATGAAGACG	AGGGGGCGCA	AAATGATATG	CCGAAAATG
	251	CCGCCGATAC	AGATAGTTG	ACACCGAATC	ACACCCGGC	TTCGAATATG
	301	CCGGCGGAA	ATATGGAAA	CCAAGCACCG	GATGCCGGGG	AATCGGAGCA
	351	GCCGGCAAC	CAACCGGATA	TGGCAAATAC	GGCGGACGGA	ATGCAGGGTG
	401	ACGATCCGTC	GGCAGGGGG	GAAAATGCCG	GCAATACGGC	TGCCCAAGGT
35	451	ACAAATCAAG	CCGAAACCAA	CTAAACCGCC	GGTTCTCAA	ATCCTGCCCTC
	501	TTCAACCAAT	CCTAGCGCA	CGAATAGCGG	TGGTGAATTT	GGAAGGACGA
	551	ACGTGGCAA	TTCTGTTGT	ATTGACGGGC	CGTCGCAAA	TATAACGTTG
	601	ACCCACTGTA	AAGGCATT	TTGTAGTGGC	AATAATTCT	TGGATGAAGA
	651	AGTACAGCTA	AAATCAGAA	TTGAAAAATT	AACTGATGCA	GACAAAATAA
40	701	GTAATTACAA	GAAAGATGGG	AAGAATGACG	GGAAGAATGA	TAAATTGTC
	751	GGTTTGGTTG	CCGATAGTGT	GCAGATGAAAG	GAATCAATC	AATATATTAT
	801	CTTTTATAAA	CCTAAACCCA	CTTCATTGTC	GCGATTTAGG	CGTTCTGCAC
	851	GGTCGAGGCG	GTCGCTTCCG	GCCGAGATGC	CGCTGATTCC	CGTCATATCAG
	901	GCGGATACGC	TGATTGTCGA	TGGGGAAGCG	GTCAGCCTGA	GGGGCATTG
45	951	CGGCAATATC	TTCGCGCCCG	AAGGGAATT	CCGGTATCTG	ACTTACGGGG
	1001	CGGAAAATT	GCCCCGGGA	TCGTATGCC	TCCGTGTTCA	AGGGAACCT
	1051	TCAAAAGGCG	AAATGCTCG	GGGCACGGCA	GTGTACAACG	GCGAAGTGC
	1101	GCATTTTCAT	ACGGAAAACG	GCCGTCCGTC	CCCGTCCAGA	GGCAGGTTG
	1151	CCGCAAAAGT	CGATTTCGGC	AGCAAATCTG	TGGACGGCAT	TATCGACAGC
50	1201	GGCGATGGTT	TGCATATGGG	TACGCAAAA	TTCAAAGCCG	CCATCGATGG
	1251	AAACGGCTTT	AAGGGGACT	GGACGGAAA	TGGCGGCCGG	GATGTTTCCG
	1301	GAAAGTTTA	CGGCCCGGCC	GGCGAGGAAG	TGGCGGGAAA	ATACAGCTAT
	1351	CGGCGAACAG	ATGCGGAAA	GGGCGGATTG	GGCGTGTGTTG	CCGGCAAA
	1401	AGAGCAGGAT	GGATCCGGAG	GAGGAGGAGC	CACAAACGAC	GACGATGTTA
55	1451	AAAAAGCTGC	CACTGTGCC	ATTGCTGCTG	CCTACACAA	TGGCCAAGAA
	1501	ATCAACGGTT	TCAAAGCTGG	AGAGACCATC	TACGACATTG	ATGAAGACGG
	1551	CACAATTACC	AAAAAAGACG	CAACTGCAGC	CGATGTTGAA	GCCGACGACT
	1601	TTAAAGGTCT	GGGTCTGAAA	AAAGTCGTGA	CTAACCTGAC	AAAACCGTC
	1651	AATGAAAACA	AACAAACGT	CGATGCCAA	GTAAAAGCTG	CAGAATCTGA
60	1701	AATAGAAAAG	TTAACAAACCA	AGTTAGCAGA	CACTGATGCC	GCTTTAGCAG
	1751	ATACTGATGC	CGCTCTGGAT	GCAACCACCA	ACGCCTTGAA	TAAATTGGGA
	1801	GAAAATATAA	CGACATTGTC	TGAAGAGACT	AAGACAAATA	TCGTAAAAT
	1851	TGATGAAAAA	TTAGAAGCCG	TGGCTGATAC	CGTCGACAAG	CATGCCAAG
	1901	CATTCAACGA	TATCGCCGAT	TCATGGATG	AAACCAACAC	TAAGGCAGAC
	1951	GAAGCGTCA	AAACCGCCAA	TGAAGCCAA	CAGACGGCCG	AAGAAACCAA
	2001	ACAAAACGTC	GATGCCAAAG	TAAAAGCTGC	AGAAACTGCGA	GCAGGCAAAG
65	2051	CCGAAGCTGC	CGCTGGCACA	GCTAATACTG	CAGCCGACAA	GGCCGAAGCT
	2101	GTCGCTGCAA	AAGTTACCGA	CATCAAAGCT	GATATCGCTA	CGAACAAAGA

2151 TAATATTGCT AAAAAAGCAA ACAGTGCCGA CGTGTACACC AGAGAAGAGT
 2201 CTGACAGCAA ATTTGTCAGA ATTGATGGTC TGAACGCTAC TACCGAAAAA
 2251 TTGGACACAC GCTTGGCTTC TGCTGAAAAA TCCATTGCCG ATCACGATAC
 2301 TCGCTGAAC GGTTTGGATA AAACAGTGTG AGACCTGCGC AAAGAAACCC
 5 2351 GCCAAGGCCT TGCAGAACAA GCCCGCTCT CCGGTCTGTT CCAACCTTAC
 2401 AACGTGGGTC GGTTCAATGT AACGGCTGCA GTCCGGCGGCT ACAAAATCCGA
 2451 ATCGGCAGTC GCCATCGGT CCGGCTTCCG CTTTACCGAA AACTTTGGCCG
 2501 CCAAGCAGG CGTGGCAGTC GGCACCTCGT CCGGTTCTTC CGCAGCCTAC
 10 2551 CATGTCGGCG TCAATTACGA GTGGTAAAAG CTT
 1 MASPDVKSAD TLSKPAAPVV SEKETEAKED APQAGSQGQG APSAQGGQDM
 51 AAVSEENTGN GGAAATDKPK NEDEGAQNDM PQNAADTDSL TPNHTPASN
 101 PAGNMENQAP DAGESEQPAN QPDmantADG MQGDDPSAGG ENAGNTAAQG
 15 151 TNQAEENNQAP GSQNPASSN PSATNSGGDF GRTNVGNSVV IDGPSQNTL
 201 THCKGDSCSG NNFLDEEVQL KSEFEKLSDA DKISNYKKDG KNDGKNDKFV
 251 GLVADSVQMK GINQYIIFYK PKPTSFARFR RSARSRRSLP AEMPLIPVNQ
 301 ADTTLVDGEA VSLTGHSIGNI FAPEGNYRYL TYGAEKLPGG SYALRVQGEP
 351 SKGEMLAGTA VYNGEVLHFF TENGRRPSPSR GRFAAKVDFG SKSVDGIIIDS
 401 GDGLHMGQTQK FKAAIDGNGF KGTWTENGGG DVSGKFYGPV GEEVAGKYSY
 451 RPTDAEKGGF GVFAGKKEQD GSggggatnd DDVKAATVA IAAAYNNGQE
 501 INGFKAGETI YDIDEDGTIT KKDATAADVE ADDFKGLIK KVVTNLTKTV
 551 NENKQNVDAK VKAAESEIEK LTTKLADTDA ALADTDAALD ATTNALNKL
 601 ENITTTFAEET KTNIVKIDEK LEAVADTVDK HAEAFNDIAD SLDETMNTKAD
 25 651 EAVKTANEAK QTAEETKQNV DAKVKAETA AGKAEEAAAGT ANTAADKA
 701 VAAKVTDIKA DIATNKDNIA KKANSADVYT REESDSKFVVR IDGLNATTEK
 751 LDTRLASAEEK SIADHDTRLN GLDKTVSDLR KETRQGLAEQ AALSGLFQPY
 801 NVGRFNVTAA VGGYKSESAV AIGTGFRFTE NFAAKAGVAV GTSSGSSAAY
 851 HVGVNYEW*

30 *ΔG983 and hybrids*

Bactericidal titres generated in response to *ΔG983* (His-fusion) were measured against various strains, including the homologous 2996 strain:

	2996	NGH38	BZ133
ΔG983	512	128	128

ΔG983 was also expressed as a hybrid, with ORF46.1, 741, 961 or 961c at its C-terminus:

ΔG983-ORF46.1

35 1 ATGACTTCTG CGCCCGACTT CAATGCAGGC GGTACCGGTA TCGGCAGCAA
 51 CAGCAGAGCA ACAACAGCGA AATCAGCAGC AGTATCTTAC GCGGGTATCA
 101 AGAACGAAAT GTGCAAAGAC AGAACGATGC TCTGTGCCGG TCAGGGATGAC
 151 GTTGCAGGTTA CAGACAGGGG TGCCAAAATC AATGCCCCCCC CCCCCGAATCT
 201 GCATACCGGA GACTTTCCAA ACCCAAATGA CGCATACAAAG AATTTGATCA
 251 ACCTCAAACC TGCAATTGAA GCAGGCTATA CAGGACGCAGG GGTAGAGGTA
 301 GGTATCGTCG ACACAGGCGA ATCCGTCGGC AGCATACTC TTCCCGAACT
 351 GTATGGCAGA AAAGAACACG GCTATAACGA AAATTACAAA AACTATACGG
 401 CGTATATGCG GAAGGAAGCG CCTGAAGACG GAGGGCGGTAA AGACATTGAA
 451 GCTTCTTCG ACGTAGAGGC CGTTATAGAG ACTGAAGCAGA AGCCGACCGA
 501 TATCCGCCAC GTAAAAGAAA TCGGACACAT CGATTTGGTC TCCCATATTA
 551 TTGGCGGGCG TTCCGTGGAC GGCAGACCTG CAGGGCGGTAT TCGGCCCGAT
 601 GCGACGCTAC ACATAATGAA TACGAATGAT GAAACCAAGA ACGAAATGAT
 651 GGTGCAGCC ATCCGCAATG CATGGGTCAA GCTGGGGCAA CGTGGCGTGC
 701 GCATCGTCAA TAACAGTTT GGAACAAACAT CGAGGGCAGG CACTGCCGAC
 751 CTTTTCCAAA TAGCCAATTG GGAGGAGCAG TACCGCAAG CGTTGCTCGA
 801 CTATTCCGGC GGTGATAAAA CAGACGAGGG TATCCGCTG ATGCAACAGA
 851 GCGATTCAGG CAACCTGTCC TACCAACATC GTAATAAAAA CATGCTTTTC
 901 ATCTTTTCGA CAGGCAATGA CGCACAAAGCT CAGCCCAACA CATATGCCCT
 951 ATTGCCATTG TATGAAAAAG ACGCTAAAAA AGGCATTATC ACAGTCGCAG
 55 1001 GCGTAGACCG CAGTGGAGAA AAGTCAAAC GGGAAATGTA TGGAGAACCG
 1051 GGTACAGAAC CGCTTGAGTA TGGCTCCAAC CATTGCGGAA TTACTGCCAT
 1101 GTGGTGCCTG TCGGCACCCCT ATGAAGCAAG CGTCCGTTTC ACCCGTACAA

5	1151	ACCCGATTCA	AATTGCCGGA	ACATCCTTTT	CCGCACCCAT	CGTAACCGGC
	1201	ACGGCGGCTC	TGCTGCTGCA	GAAATACCCG	TGGATGAGCA	ACGACAAACCT
	1251	GCGTACACG	TTGCTGACGA	CGGCTCAGGA	CATCGGTGCA	GTCGGCGTGG
	1301	ACAGCAAGTT	CGGCTGGGGA	CTGCTGGATG	CGGGTAAGGC	CATGAACCGGA
10	1351	CCCGCGTCCT	TTCCGTTCGG	CGACTTTACG	GCCGATACGA	AAGGTACATC
	1401	CGATATTGCC	TACTCCTTCC	GTAACGACAT	TTCAGGCACG	GGCGGCGTGA
	1451	TCAAAAAAGG	CGGCAGCCAA	CTGCAACTGC	ACGGCAACAA	CACCTATACG
	1501	GGCAAAACCA	TTATCGAAGG	CGGTCGCTG	GTGTTGTACG	GCAACAAACAA
15	1551	ATCGGATATG	CGCGTCGAAA	CCAAAGGTGC	GCTGATTTAT	AACGGGGCGG
	1601	CATCCGGCGG	CAGCCTGAAC	AGCGACGGCA	TTGTCATCT	GGCAGATACC
	1651	GACCAATCCG	GCGAAACGA	AACCGTACAC	ATCAAAGGCA	GTCTGCAGCT
	1701	GGACGGCAA	GGTACGCTGT	ACACACGTTT	GGGCAAACCTG	CTGAAAGTGG
20	1751	ACGGTACGGC	GATTATCGGC	GGCAAGCTGT	ACATGTCGGC	ACGGCGCAAG
	1801	GGGGCAGGCT	ATCTCAACAG	TACCGGACGA	CGTGTCCCT	TCCTGAGTGC
	1851	CGCCAAATAC	GGGCAGGATT	ATTCTTCTC	CACAAACATC	GAAACCGACG
	1901	GGCGCTGTCT	GGCTTCCCTC	GACAGCGTCG	AAAAAACAGC	GGGCAGTGAA
	1951	GGGCACACGC	TGTCTTATTA	TGTCCGTGCG	GGCAATGCGG	CACGGACTGC
25	2001	TTCGGCAGCG	GCACATTCCG	CGCCCGCCGG	TCTGAAACAC	GCCGTTAGAAC
	2051	AGGGCGGCAG	CAATCTGGAA	AACCTGATGG	TGAACTTGGG	TGCCTCCGAA
	2101	TCATCCGCAA	CACCCGAGAC	GGTTGAAACT	GGGGCAGCCG	ACCGCACAGA
	2151	TATGCCGGC	ATCCGCCCT	ACGGCGAAC	TTTCCGCGCA	GGGGCAGCCG
	2201	TACACCATGC	GAATGCCGCC	GACGGTGTAC	GCATCTTCAA	CAGTCTCGCC
	2251	GCTACCGTCT	ATGCCGACAG	TACCGCCGCC	CATGCCGATA	TGCAGGGACG
30	2301	CCGCTTAA	GCCGTATCGG	ACGGGTTGGG	CCACAAACGGC	ACGGGCTCTGC
	2351	CGCTCATCGC	GCAAAACCAA	CAGGACGGTG	GAACGTGGGA	ACAGGGCGGT
	2401	GTGAAAGCA	AAATGCGGG	CAGTACCCAA	ACCGTGGCA	TTGCGCGAA
	2451	AACCGGCAGA	AATACGACAG	CAGCCGCCAC	ACTGGGCATG	GGACGCCAGCA
	2501	CATGGAGCGA	AAACAGTGC	AATGCAAAA	CCGACAGCAT	TAGTCTGTTT
	2551	GCAGGCATAC	GGCACGATGC	GGGGGATATC	GGCTATCTCA	AAGGCCTGTT
35	2601	CTCCTACGGA	CGCTACAAAA	ACAGCATTAG	CCGCAGCACC	GGTGGGACG
	2651	AAACATGCCG	AGGCAGCGTC	AAACGGCACGC	TGATGCCAGCT	GGGGCCACTG
	2701	GGCGGTGTCA	ACGTTCCGTT	TGCGCAACG	GGAGATTTGA	CGGTGAAAGG
	2751	CGGCTCGCG	TACGACCTGC	TCAAACAGGA	TGCATTGCGC	GAAAAGGCA
	2801	GTGCTTGGG	CTGGAGGGC	AACAGCCTCA	CTGAAGGCAC	GCTGGTGGGA
40	2851	CTCGGGGTC	TGAAGTGTG	GCAACCCCTG	AGCGATAAAG	CCGTCTGTT
	2901	TGCAACGGCG	GGCGTGGAAAC	GCGACCTGAA	CGGACGCGAC	TACACGGTAA
	2951	CGGGCGGCTT	TACCGGGCG	ACTGCAGCAA	CCGGCAAGAC	GGGGGCACGC
	3001	AATATGCCG	ACACCCGCT	GGTTGCCGGC	CTGGGCGCGG	ATGTGAAATT
	3051	CGGCAACGGC	TGGAACGGCT	TGGCACGTTA	CAGCTACGCC	GGTTCCAAAC
45	3101	AGTACGGCAA	CCACAGCGGA	CGAGTCGGCG	TAGGCTACCG	GTTCCCTGAC
	3151	GGTGGCGGAG	GCACTGGATC	CTCAGATTTG	GCAAAACGATT	CTTTTATCCG
	3201	GCAGGTTCTC	GACCGTCAGC	ATTTCGAACC	CGACGGGAAA	TACCACTTAT
	3251	TCGGCAGCAG	GGGGGAACCT	GCCGAGCGCA	GGGGCCATAT	CGGATTGGGA
50	3301	AAAATACAAA	GCCATCAGTT	GGGCAACCTG	ATGATTCAAC	AGGCGCCAT
	3351	TAAGGAAAT	ATCGGCTACA	TTGTCGCGTT	TTCCGATCAC	GGGCACGAAG
	3401	TCCATTCCCC	CTTCGACAA	CATGCCCTAC	ATTCCGATTC	TGATGAAGCC
	3451	GGTAGTCCCG	TTGACGGATT	TAGCCTTAC	CGCATCCATT	GGGACGGATA
	3501	CGAACACCAT	CCCGCCGACG	GCTATGACGG	GCCACAGGGC	GGCGGCTATC
55	3551	CCGCTCCCAA	AGGCGCGAGG	GATATATACA	GCTACGACAT	AAAAGGCATT
	3601	GCCAAATAA	TCCGCTCAA	CCTGACCGAC	AACCGCAGCA	CCGGACAAACG
	3651	GCTTGCAC	CGTTCCAC	ATGCGGGTAG	TATGCTGACG	CAAGGAGTAG
	3701	GCGACGGATT	CAAACGCC	ACCCGATACA	CCCCCGAGCT	GGACAGATCG
	3751	GGCAATGCCG	CCGAAGCCTT	CAACGGCACT	CGAGATATCG	TTAAAAACAT
60	3801	CATCGGCGCG	GCAGGAGAAA	TTGTCGGCGC	AGGGCATGCC	GTGCAGGGCA
	3851	TAAGCGAAGG	CTCAAACATT	GCTGTCACTG	ACGGCTTGGG	TCTGCTTCC
	3901	ACCGAAAACA	AGATGGCGCG	CATCAACGAT	TTGGCAGATA	TGGCGCAACT
	3951	CAAAGACTAT	GCCGCAGCAG	CCATCCGCGA	TTGGGCAGTC	AAAAACCCCA
	4001	ATGCCGCACA	AGGCATAGAA	GCCGTCAGCA	ATATCTTAT	GGCAGCCATC
	4051	CCCACCAAAG	GGATTGGAGC	TGTTCGGGGA	AAATACGGCT	TGGGCGGCAT
	4101	CACGGCACAT	CCTATCAAGC	GGTCGCAGAT	GGGCGCGATC	GCATTGCGA
	4151	AAGGGAAATC	CGCCGTCAGC	GACAATTTG	CCGATGCGGG	ATACGCCAAA
	4201	TACCGTCCC	CTTACCATTC	CCGAAATATC	CCTTCAAAC	TGGAGCAGCG
	4251	TTACGGCAA	GAAAACATCA	CCTCCTCAAC	CCTGCCCG	TCAAACGGCA
65	4301	AAAATGTCAA	ACTGGCAGAC	CAACGCCACC	CGAAGACAGG	CGTACCGTT
	4351	GACGGTAAAG	GGTTTCCGAA	TTTGAGAAG	CACGTAAAT	ATGATACGCT
	4401	CGAGCACCAC	CACCAACCAC	ACTGA		

1	MTSAPDFNAG	GTGIGGSNSRA	TTAKSAAVSY	AGIKNEMCKD	RSMLCAGRDD
51	VAVTDRDAKI	NAPPPNLHTG	DFPNPNDAYK	NLINLKPATE	AGYTGRGVEV
101	GIVDTGESVG	SISFPELYGR	KEHGYNENYK	NYTAYMRKEA	PEDGGGK DIE
151	ASFDEAVIE	TEAKPTDIRH	VKEIGHIDL V	SHIIGGRSVD	GRPAGGIAPD
5	201	ATLHIMNTND	ETKNEMMVA	IRNAWVKLG E	RGVRIVNNSF
	251	LFOQIANSEEQ	YRQALLDYS G	GDKTDEGIRL	MQQSDYGNL S
	301	IFSTGNDQAQ A	QPNTYALLPF	YEKDAQKGII	TVAGVDRSGE
	351	GTEPLEYGSN	HCGITAMWCL	SAPYEASVRF	KFKREMYGEP
10	401	TAALLLQKYP	WMSNDNLRTT	LLTTAQDIGA	TSFSAPIVTG
	451	PASFPFGDFT	ADTKGTSDIA	Y SFRNDISGT	GGLIKGGGSQ
	501	GKTIIEGGSL	VLYGNNKSDM	RVETKGALIY	LQLHGNNTYT
	551	DQSGANETVH	IKGSLQLDGK	GTLYTRLGKL	LDAGKAMNG
	601	GAGYLNSTGR	RVPFLSAKI	GQDYSFFTNI	SDGIVYLADT
15	651	GDTLSYYVRR	GNAARTASAA	AHSAPAGLKH	GKLYMSARGK
	701	SSATPETVET	AAADRTDMPG	IRPYGATFRA	AVEQGGSNLE
	751	ATVYADSTAA	HADMQGRRLK	AVSDGLDHNG	NLMVELDASE
	801	VEGKMRGSTQ	TVGIAAKTGE	NTTAAATLGM	GRSTWSENSA
	851	AGIRHDAGDI	GYLKGLFSY G	RYKNSISRST	NAKTDISL F
20	901	GGVNPVFAAT	GDLTVEGGLR	YDLIKQDAFA	GGVGYRFLD
	951	LAGLKL SQPL	SDKAVLFATA	GVERDLNGRD	YTVTGGFTGA
	1001	NMPHTRLVAG	LGADVEFGNG	WG LARYSYA	TAATGKTGAR
	1051	GGGGTGSSDL	ANDSFIRQL	DRQHFEPDGE	NGTLMQLGAL
	1101	KIQSHQLGNL	MIQQAAIKGN	YHLFGSRGEL	AERSHIGLG
25	1151	GSPVDFGSLY	RIHWDGYEH H	PADGYDGPQG	GHEVHSPFDN
	1201	AQNIRLNLT D	NRSTGQRLAD	RFHNAGSMLT	HASHSDSDEA
	1251	GNAAEAFNGT	ADIVKNIIGA	AGEIVGAGDA	DIYSYDIKGV
	1301	TENKMARIND	LADMAQLKD Y	AAA AIRDWA V	TRYSP ELD RS
	1351	PIKGIGAVRG	KYGLGGITAH	PIKRSQMGAI	AVSNIFMAAI
30	1401	YPSPYHSRNI	RSNLEQRYGK	ENITSSTVPP	DNFADAAYAK
	1451	DGKGFPNFEK	HVKYDTLEHH	HHHH*	QRHPKTGVPF

ΔG983-741

35	1	ATGACTTCTG	CGCCCGACTT	CAATGCAGGC	GGTACCGGTA
	51	CAGCAGAGCA	ACAACAGCGA	AATCAGCAGC	AGTATCTTAC
	101	ACAACGAAAT	GTGCAAAGAC	AGAACGATGC	GCCGGTATCA
	151	GTTGCGGTTA	CAGACAGGGA	TGCCAAAATC	TCTGTGCCGG
	201	GCATACCGGA	GACTTTCCAA	ACCCAAATGA	AATGCCCT
40	251	ACCTCAAACC	TGCAATTGAA	GCAGGCTATA	GGTAGAGGTA
	301	GGTATCGTCG	ACACAGGCGA	ATCCGTCGGC	AGCATATCCT
	351	GTATGGCAGA	AAAGAACACG	GCTATAAACG A	TTCCCGAACT
	401	CGTATATGCG	GAAGGAAGCG	CCTGAAGACG	AACTATACGG
	451	GCTCTTTTCG	ACGATGAGGC	CGTTATAGAG	AGACATTGAA
45	501	TATCCGCCAC	GTAAAAGAAA	TCGGACACAT	AGCCGACGGA
	551	TTGGCGGGCG	TTCCGTGGAC	GGCAGACCTG	CGATTGGTC
	601	GCGACGCTAC	ACATAATGAA	TACGAATGAT	TCCCATATT A
	651	GGTTGCAGCC	ATCCGCAATG	CATGGTCAA	TGCGCCGAT
50	701	GCATCGTCAA	TAACAGTTT	GGAACAACAT	CGAGGGCAGG
	751	CTTTTCCAAA	TAGCCAATTC	GGAGGGAGCAG	CACTGCCGAC
	801	CTATTCCGGC	GGTGATAAAA	CAGACGAGGG	TACGCCAAG
	851	GCGATTACCG	CAACCTGTCC	TACCACATCC	CGTTGCTCGA
	901	ATCTTTTCG A	CAGGCAATGA	CGCACAAAGCT	ATGCAACAGA
	951	ATTGCCATT T	TATGAAAAG	ACGCTCAAAA	CATATGCCCT
55	1001	GCGTAGACCG	CAGTGGAGAA	AAAGTCAAAC	AGGCATTATC
	1051	GGTACAGAAC	CGCTTGAGTA	CATTGCGGAA	ACAGTCGCAG
	1101	GTGGTGCCTG	TCGGCACCT	ATGAAGCAAG	TTACTGCCAT
	1151	ACCCGATTCA	AATTGCCGG A	ACATCCTTT	ACCCGTACAA
60	1201	ACGGCGGCTC	TGCTGCTGCA	GAAATACCG	CGTAACCGGC
	1251	GCGTACCA CG	TTGCTGACGA	CGGCTCAGGA	TGGATGAGCA
	1301	ACAGCAAGTT	CGGCTGGGG A	CTGCTGGATG	ACGACAACCT
	1351	CCCGCGTCCT	TTCCGTTCGG	CGACTTTACC	GTCGGCGTGG
	1401	CGATATTGCC	TACTCCTTCC	GTAAACGACAT	AAGGTACATC
	1451	TCAAAAAGG	CGGCAGCAA	CTGCAACTGC	GGCGGCCTGA
	1501	GGCAAAACCA	TTATCGAAGG	CGGTTCGCTG	ACGGCAACAA
65	1551	ATCGGATATG	CGCGTCGAAA	CACAGGTGC	CACCTATAACG
	1601	CATCCGGCGG	CAGCCTGAAC	AGCGACGGCA	GTCAGATACC
	1651	GACCAATCCG	GCGAAACGA	AACCGTACAC	GGCAGATA CC

1701	GGACGGAAA	GGTACGCTGT	ACACACGTTT	GGGCAAACGT	CTGAAAGTGG
1751	ACGGTACGGC	GATTATCGGC	GGCAAGCTGT	ACATGTCGGC	ACGCGGCAAG
1801	GGGGCAGGCT	ATCTCAACAG	TACCGGACGA	CGTGTTCCTC	TCCTGAGTGC
1851	CGCCAAAATC	GGGCAGGATT	ATTCTTCTT	CACAAACATC	GAAACCGACG
1901	GCGGCCTGCT	GGCTTCCCTC	GACAGCGTCG	AAAAAACAGC	GGGCAGTGAA
1951	GGCGACACGC	TGTCCCTATTA	TGTCCGTCGC	GGCAATGCGG	CACGGACTGC
2001	TTCGGCAGCG	GCACATTCGG	CGCCCGCCGG	TCTGAAACAC	GCCGTAGAAC
2051	AGGGCGGCAG	CAATCTGGAA	AACCTGATGG	TCGAAGCTGGA	TGCCTCCGAA
2101	TCATCCGCAA	CACCCGAGAC	GGTTGAAAAT	GCAGGAGCCG	ACCGCACAGA
2151	TATGCCGGGC	ATCCGCCCT	ACGGCGAAC	TTTCCGCGCA	GCAGGAGCCG
2201	TACAGCATGC	GAATGCCGCC	GACGGTGTAC	GCATCTTCAA	CAGTCTCGCC
2251	GCTACCGTCT	ATGCCGACAG	TACCGCCGCC	CATGCCGATA	TGCAGGGACG
2301	CCGCTGAAA	GCGTATCGG	ACGGGTTGGA	CCACAAACGGC	ACGGGTCTGC
2351	GCGTCATCGC	GCAAAACCAA	CAGGACGGTG	GAACGTGGGA	ACAGGGCGGT
2401	GTTGAAGGCA	AAATGCGCGG	CAGTACCCAA	ACCGTCGGCA	TTGCCCGGAA
2451	AACCGGGCAA	AATACGACAG	CAGCCGCCAC	ACTGGGCATG	GGACGCAGCA
2501	CATGGAGCGA	AAACAGTGC	AATGCAAAA	CCGACAGCAT	TAGTCTGTTT
2551	GCAGGCATAC	GGCACGATGC	GGGGGATATC	GGCTATCTCA	AAGGCCTGTT
2601	CTCCTACGGA	CGCTACAAAA	ACAGCATCAG	CCGCAGCACC	GGTGCGGACG
2651	AAACATGCCGA	AGGCAGCGTC	AACGGCACGC	TGATGCGAGCT	GGGCGCACTG
2701	GGCGGTGTCA	ACGTTCCGTT	TGCCGCAACG	GGAGATTTGA	CGGTCGAAGG
2751	CGGTCTGC	TACGACCTGC	TCAAACAGGA	TGCATTGCC	GAAAAAAGGCA
2801	GTGCTTTGGG	CTGGAGCGGC	AACAGCCTCA	CTGAAGGCCAC	GCTGGTCGGA
2851	CTCGGGGTC	TGAAGCTGTC	GCAACCCCTTG	AGCGATAAAG	CCGTCTGTT
2901	TGCAACGGCG	GGCGTGGAAAC	GCGACCTGAA	CGGACGCGAC	TACACGGTAA
2951	CGGGCGGCTT	TACCGGGCG	ACTGCAGCAA	CCGGCAAGAC	GGGGGCACGC
3001	AATATGCCGC	ACACCCGTCT	GGTTGCCGGC	CTGGGCGGG	ATGTCGAATT
3051	CGGCAACGGC	TGGAACGGCT	TGGCACGTTA	CAGCTACGCC	GGTTCCAAC
3101	AGTACGGCAA	CCACAGCGGA	CGAGTCGGCG	TAGGCTACCG	GTTCCCTCGAG
3151	GGATCCGGAG	GGGGTGGTGT	CGCCGCGAC	ATCGGTGCGG	GGCTTGCCGA
3201	TGCACTAAC	GCACCGCTCG	ACCATAAAGA	CAAAGGTTG	CAGTCCTTGA
3251	CGCTGGATCA	GTCCGTCAGG	AAAACAGGAGA	AACTGAAGCT	GGCGGGCACAA
3301	GGTGCAGGAA	AAACTTATGG	AAACGGTGCAC	AGCCTCAATA	CGGGCAAATT
3351	GAAGAACGAC	AAGGTAGGCC	GTTTCGACTT	TATCGCCCAA	ATCGAAGTGG
3401	ACGGGCAGCT	CATTACCTTG	GAGAGTGGAG	AGTTCCAAGT	ATACAAACAA
3451	AGCCATTCCG	CCTTAACCGC	CTTTCAGACC	GAGCAAATAC	AAGATTCGGA
3501	GCATTCCGGG	AAGATGGTTG	CGAAACGCCA	GTTCAGAAC	GGCGACATAG
3551	CGGGCGAAC	TACATCTTT	GACAAGCTTC	CCGAAGGCCG	CAGGGCGACA
3601	TATCGCGGG	CGGCGTTCGG	TTCAGACGAT	GCCGGCGGAA	AACTGACCTA
3651	CACCATAGAT	TTCGCCGCCA	AGCAGGGAAA	CGGCAAACATC	GAACATTG
3701	AAATGCCAGA	ACTCAATGTC	GACCTGGCCG	CCGCGCATAT	CAAGCCGGAT
3751	GGAAAACGCC	ATGCCGTCA	CAGCGGTTCC	GTCTTTACA	ACCAAGCCGA
3801	GAAAGGCAGT	TACTCCCTCG	GTATCTTGG	CGGAAAAGCC	CAGGAAGTTG
3851	CCGGCAGCGC	GGAAAGTGA	ACCGTAAACG	GCATACGCCA	TATCGGCCTT
3901	GCCGCCAAC	AACTCGAGCA	CCACCAACCAC	CACCACTGA	
1	MTSAPDFNAG	GTGIGNSNSA	TTAKSAAVSY	AGIKNEMCKD	RSMLCAGRDD
51	VAVTDRDAKI	NAPPNLHTG	DFPNPNDAYK	NLNLNLPKPAIE	AGYTGRGVEV
101	GIVDTGESVG	SISFPELYGR	KEHGYNENYK	NYTAYMRKIEA	PEDGGGKDIE
151	ASFDEEAVIE	TEAKPTDIRH	VKEIGHIDLV	SHIIIGGRSVD	GRPAGGIAPD
201	ATLHIMNTND	ETKNEMMVA	IRNAWVKLGE	RGVRIVNNNSF	GTTSRAGTAD
251	LFQIANSEEQ	YRQALLDYS	GDKTDEGIRL	MQQSDYGNLS	YHIRKNMLF
301	IFSTGNDQAQ	QPNTYALLPF	YEKDAQKII	TVAGVDRSCE	KFKREMYGEP
351	GTEPLEYGSN	HCGITAMWCL	SAPYEASVRF	TRTNPIQIAG	TSFSAPIVTG
401	TAALLLQKYP	WMSNDNLRTT	LLTTAQDIGA	VGVDLSKFGWG	LLDAGKAMNG
451	PASFPFGDFT	ADTKGTSIA	YSFRNDISGT	GGLIKKGSQ	LQLHGNNTYT
501	GKTIIEGGSL	VLYGNNKSDM	RVETKGALIY	NGAASGGSILN	SDGIVYLADT
551	DQSGANETVH	IKGSLQLDGK	GTLYTRLGKL	LKDVGTAIIG	GKLYMSARGK
601	GAGYLNSTGR	RVPFLSAAKI	GQDYSFFTNI	ETDGGLLASL	DSVEKTAGSE
651	GDTLSYYVRR	GNAARTASAA	AHSAPAGLKH	AVEQGGSNLE	NLMVELDASE
701	SSATPETVET	AAADRTDMPG	IRPYGATFRA	AAAQVHANAA	DGVRIFNSLA
751	ATVYADSTAA	HADMQGRRLK	AVSDGLDHNG	TGLRVIAQTQ	QDGTTWEQGG
801	VEGKMRGSTQ	TVGIAAKTGE	NTTAAATLGM	GRSTWSENSA	NAKTDSISLF
851	AGIRHDAGDI	GYLKGLFSYG	RYKNSISRST	GADEHAEGSV	NGTLMQLGAL
901	GGVNVPFAAT	GDLTVEGGLR	YDLLKQDFA	EKGSLGWG	NSLTEGTLVG
951	LAGLKLSQPL	SDKAVLFATA	GVERDLNGRD	YTVTGGFTGA	TAATGKTGAR
1001	NMPHTRLVAG	LGADVEFGNG	WNGLARYSYA	GSKQYGNHSG	RVGVGYRFLE

1051 GSGGGGVAAD IGAGLADALT APLDHDKGL QSLTLDQSVR KNEKLKLAQ
 1101 GAEKTYGNGD SLNTGKLKD KVSRDFD FIRQ IEVDGQLITL ESGEFQVYKQ
 1151 SHSALTAFQQT EQIQDSEHSG KMVAKRQFRI GDIAGEHTSF DKLPEGGRAT
 5 1201 YRGTAFGSDD AGGKLTYTID FAAKQGNGKI EHLKSPELNV DLAAADIKPD
 1251 GKRHAVISGS VLYNQAEKGS YSLGIFGGKA QEVAGSAEVK TVNGIRHIGL
 1301 AAKQLEHHHH HH*

ΔG983-961

10 1 ATGACTTCTG CGCCCGACTT CAATGCAGGC GGTACCGGTA TCGGCAGCAA
 51 CAGCAGAGCA ACAACAGCGA AATCAGCAGC AGTATCTTAC GCCGGTATCA
 101 AGAACGAAAT GTGCAAAGAC AGAACATGC TCTGTGCCGG TCGGGATGAC
 15 151 GTTGCCTTAA CAGACAGGGA TGCCAAAATGAT ATGCCCTTAC CCCCGAATCT
 201 GCATACCGGA GACTTTCCAA ACCCAAATGAT CGCATAACAAG AATTGATCA
 251 ACCTCAAACC TGCAATTGAA GCAGGCTATA CAGGACGCGG GGTAGAGGTA
 301 GGTATCGTCG ACACAGGCGA ATCCGTCGGC AGCATATCCT TTCCCGAACT
 351 GTATGGCAGA AAAGAACACG GCTATAACGA AAATTACAAA AACTATACGG
 401 CGTATATGCG GAAGGAAGCG CCTGAAGACG GAGGCGGTAA AGACATTGAA
 451 GCTCTTTCTG ACGATGAGGC CGTTATAGAG ACTGAAGCAA AGCCGACGGA
 20 501 TATCCGCCAC GTAAAAGAAA TCGGACACAT CGATTTGGTC TCCCATATTA
 551 TTGGCGGGCG TTCCGTGGAC GGCAGACCTG CAGGCGGTAT TCGGCCGAT
 601 GCGACGCTAC ACATAATGAA TACGAATGAT GAAACCAAGA ACGAAATGAT
 651 GGTGCGAGC ATCCGCAATGAT CATGGTCAA GCTGGGCGAA CGTGGCGTGC
 701 GCATCGTCAA TAACAGTTT GGAACAAACAT CGAGGGCAGG CACTGCCGAC
 751 CTTTTCCAAA TAGCCAATTG GGAGGAGCAG TACCGCCAAG CGTTGCTCGA
 801 CTATTCCGGC GGTGATAAAA CAGACGAGGG TATCCGCTG ATGCAACAGA
 851 GCGATTACGG CAACCTGTCC TACACACATCC GTAATAAAAAA CATGCTTTTC
 901 ATCTTTTCGA CAGGCAATGAT CGCACAAGCT CAGCCCAACA CATATGCCCT
 951 ATTGCCATT TATGAAAAAG ACGCTAAAAA AGGCATTATC ACAGTCGCG
 30 1001 GCGTAGACCG CAGTGGAGAA AAGTCAAAC GGGAAATGTA TGGAGAACCG
 1051 GGTACAGAAC CGCTTGAGTA TGGCTCCAAC CATTGCGGAA TTACTGCCAT
 1101 GTGGTGCCTG TCGGCACCCAT ATGAAGCAAG CGTCCGTTTC ACCCGTACAA
 1151 ACCCGATTCA AATTGCCGA ACATCCCTTT CGCGACCCAT CGTAACCGGC
 1201 ACGCGGGCTC TGCTGCTGCA GAAATACCCG TGGATGAGCA ACGACAACCT
 35 1251 GCGTACCAAG TTGCTGACGA CGGCTCAGGA CATCGGTGCA GTCGGCGTGG
 1301 ACAGCAAGTT CGGCTGGGAA CTGCTGGATG CGGGTAAGGC CATGAACGGA
 1351 CCCCGTCTCCT TTCCGTTCGG CGACTTTTACCC GCGGATACGA AAGGTACATC
 1401 CGATATTGCC TACTCCTCC GTAAACGACAT TTCAGGCACG GGCGGCCTGA
 1451 TCAAAAAAGG CGGCAGCCAA CTGCAACTGC ACGGCAACAA CACCTATACG
 40 1501 GGCAAAACCA TTATCGAAGG CGGTTGCGCTG GTGTTGTCACG GCAACAACAA
 1551 ATCGGATATG CGCGTCGAAA CCAAAGGTGC GCTGATTTAT AACGGGGCGG
 1601 CATCCGGCGG CAGCTGAAAC AGCGACGGCA TTGTCATCT GGCAGATACC
 1651 GACCAATCCG GCGCAAACGA AACCGTACAC ATCAAAGGCA GTCTGCAGCT
 1701 GGACGGCAAA GGTACGCTGT ACACACGTTT GGGCAAACGT CTGAAAGTGG
 45 1751 ACGGTACGGC GATTATCGGC GGCAAGCTGT ACATGTCGGC ACGGGCGAAG
 1801 GGGGCAGGCT ATCTCAACAG TACCGGACGA CGTGTCCCT TCCTGAGTGC
 1851 CGCCAAAATC GGGCAGGATT ATTCTTTCTT CACAAACATC GAAACCGACG
 1901 GCGGCCTGCT GGCTTCCCTC GACAGCGTCG AAAAAACAGC GGGCAGTGA
 1951 GGCACACGC TGTCCTATTA TGTCGTCGC GGCAATGCGG CACGGACTGC
 50 2001 TTGCGCAGCG GCACATTCCG CGCCCGCCGG TCTGAAACAC GCGCTAGAAC
 2051 AGGGCGGAG CAATCTGAA AACCTGATGG TCGAAGTGG TGCCCTCCGAA
 2101 TCATCCGCAA CACCCGAGAC GGTGAAACT CGGGCAGCCG ACCGCACAGA
 2151 TATGCCGGGC ATCCGCCCT ACGGCGCAAC TTTCCGCGCA GCGGCAGCCG
 2201 TACAGCATGC GAATGCCGCC GACGGTGTAC GCATCTTCAA CAGTCTCGCC
 55 2251 GCTACCGTCT ATGCCGACAG TACCGCCGCC CATGCCGATA TGCAAGGGACG
 2301 CCGCCTGAAA GCGTATCGG ACGGGTTGGA CCACAAACGGC ACGGGTCTGC
 2351 GCGTCATCGC GCAAACCCAA CAGGACGGTG GAACGTGGGA ACAGGGCGGT
 2401 GTTGAAGGCA AAATGCGCGG CAGTACCCAA ACCGTGGCA TTGCGCGAA
 2451 AACCGGCAGA AATACGACAG CAGCCGCCAC ACTGGGCATG GGACGGCAGCA
 60 2501 CATGGAGCGA AAACAGTGC AATGCAAAAAA CCGACAGCAT TAGTCTGTTT
 2551 GCAGGCATAC GGCACGATGC GGGCGATATC GGCTATCTCA AAGGCCTGTT
 2601 CTCCTACGGA CGCTACAAA ACAGCATCAG CGCGACGCC ACCGGCGACG
 2651 AACATGCGGA AGGCAGCGTC AACGGCACGC TGATGCGAGCT GGGCGCACTG
 2701 GCGGGTGTCA ACGTTCCGTT TGCGCAACG GGAGATTTGA CGGTGCAAGG
 2751 CGGTCTCGC TACGACCTGC TCAAACAGGA TGCATTGCC GAAAAGGCA
 2801 GTGCTTTGGG CTGGAGCGGC AACAGCCTCA CTGAAGGCAC GCTGGTCGGA
 2851 CTCGCGGGTC TGAAGCTGTC GCAACCCCTG AGCGATAAAAG CCGTCCTGTT

2901	TGCAACGGCG	GGCGTGGAAC	GCGACCTGAA	CGGACGCGAC	TACACGGTAA
2951	CGGGCGGCFT	TACCGGCCGCG	ACTGCAGCAA	CCGGCAAGAC	GGGGCACGC
3001	AATATGCCGC	ACACCCGTCT	GGTTGCCGGC	CTGGGCCGGG	ATGTCGAATT
3051	CGGCAACGGC	TGGAACGGCT	TGGCACGTTA	CAGCTACGCC	GGTTCCAAC
5	3101	AGTACGGCAA	CCACAGCGGA	CGAGTCGGCG	TAGGCTACCG
	3151	GGTGGCGGAG	GCACGGATC	CGCCACAAAC	GACGACGATG
	3201	TGCCACTGTG	GCCATTGCTG	CTGCTACAA	CAATGCCAA
	3251	GTTCAAAGC	TGGAGAGACC	ATCTACGACA	TTGATGAAGA
10	3301	ACCAAAAAAG	ACGCAACTGC	AGCCGATGTT	GAAGCCGACG
	3351	TCTGGGTCTG	AAAAAAGTCG	TGACTAACCT	GACCAAAACC
	3401	ACAAACAAAA	CGTCGATGCC	AAAGTAAAAG	CTGCAGAATC
	3451	AAGTTAACAA	CCAAGTTAGC	AGACACTGAT	GCCGCTTTAG
15	3501	TGCCGCTCTG	GATGCAACCA	CCAACGCCCT	GAATAAAATTG
	3551	TAACGACATT	TGCTGAAGAG	ACTAAGACAA	ATATCGTAAA
	3601	AAATTAGAAG	CCGTGGCTGA	TACCGTCGAC	AAGCATGCCG
	3651	CGATATGCC	GATTCAATTG	ATGAAACCAA	CACTAAGGCA
20	3701	TCAAAACCGC	CAATGAAGCC	AAACAGACGG	CCGAGAAC
	3751	GTCGATGCCA	AAGTAAAAGC	TGCAGAAACT	GCAGCAGGCC
	3801	TGCCGCTGGC	ACAGCTAATA	CTGCAGCCGA	CAAGGCCGAA
	3851	CAAAAGTTAC	CGACATCAAA	GCTGATATCG	CTACGAACAA
	3901	GCTAAAAAAG	CAAACAGTGC	CGACGTGTAC	ACCAGAGAAAG
	3951	CAAATTGTC	AGAATTGATG	GTCTGAACGC	TACTACCGAA
25	4001	CACGCTTGGC	TTCTGCTGAA	AAATCCATTG	CCGATCACGA
	4051	AACGGTTGG	ATAAAAACAGT	GTCAGACCTG	CGCAAAGAAA
	4101	CCTTGCAGAA	CAAGCCGCGC	TCTCCGGTCT	GTTCCAACCT
	4151	GTCGGTTCAA	TGTAACGGCT	GCAGTCGGCG	GCTACAAATC
	4201	GTCGCCATCG	GTACCGGCTT	CCGCTTTACC	AAAAACTTIG
	4251	AGGCGTGGCA	GTCGGCACCT	CGTCGGTTC	TTCCGCAGCC
30	4301	GCGTCAATT	CGAGTGGCTC	GAGCACCACC	ACCACCACTA
	1	MTSAPDFNAG	GTGIGSNSRA	TTAKSAAVSY	AGIKNEMCKD
	51	VAVTDRAKI	NAPPNLHTG	DFPNPNDAYK	NLINLKPRAIE
35	101	GIVDTGESVG	SISFPELYGR	KEHGYNENYK	NYTAYMRKEA
	151	ASFDEAVIE	TEAKPTDIRH	VKEIGHIDL	SHIIGGRSVD
	201	ATLHIMNTND	ETKNEMMVA	IRNAWVKLGE	RGVRIVNNSF
	251	LFQIANSEEQ	YRQALLDYS	GDKTDEGIRL	MQQSDYGNLIS
	301	IFSTGNDAQA	QPNTYALLPF	YEKDAQKGI	TVAGVDRSCE
	351	GTEPLEYGSN	HCGITAMWCL	SAPYEASVRF	TRTNPIQIAG
40	401	TAALLQKYP	WMSNDNLRTT	LLTQAQDIGA	VGVDSDKFGWG
	451	PASFPFGDFT	ADTKGTSIA	YSFRNDISGT	GGLIKGGSQ
	501	GKTIIEGGSL	VLYGNNSKDM	RVETKGALIY	NGAASGGSLN
	551	DQSGANETVH	IKGSLQLDGK	GTLYTRLGKL	LKVDGTTAIC
	601	GAGYLNSTGR	RVPFLSAKI	QDYSFFTNI	ETDGGLLASL
45	651	GDTLSSYVRR	GNAARTASAA	AHSAPAGLKH	AVEQGGSNLE
	701	SSATPETVET	AAADRTDMPG	IRPYGATFRA	AAAVQHANAA
	751	ATVYADSTAA	HADMQGRRLK	AVSDGLDHNG	TGLRVIAQTQ
	801	VEGKMRGSTQ	TVGIAAKTGE	NTTAAATLGM	GRSTWSENSA
	851	AGIRHDAGDI	GYLKGLFSYG	RYKNSISRST	GADEHAEGSV
50	901	GGVNPFAAT	GDLTVEGGGLR	YDLLKQDFA	EKGSLGWG
	951	LAGLKLSQLP	SDKAVLFATA	GVERDLNGRD	YTVTGGFTGA
	1001	NMPHTRLVAG	LGADVEFGNG	WNGLARYSYA	GSKQYGNHSG
	1051	GGGGTGSATN	DDDVKKAATV	AIAAAAYNNGQ	EINGFKAGET
55	1101	TKDADTAADV	EADDFKGLG	KKVVTNLTKT	VNENKQNVDA
	1151	KLTTKLADTD	AALADTDAAL	DATTNALNKL	GENITFAEE
	1201	KLEAVADTV	KHAEAFNDIA	DSLDETNTKA	TKTNIVKIDE
	1251	VDAKVKAAET	AAGKAEAAAG	TANTAADKAE	AVAAKVTDIK
	1301	AKKANSADVY	TREESDSKFV	RIDGLNATTE	KLDTRLASAE
	1351	NGLDKTVSDL	RKETRQGLAE	QAALSGLFQP	YNVGRFNVTA
60	1401	VAIGTGFRFT	ENFAAKAGVA	VGTSSGSSAA	YHVGVNYEWL
					EEHHHHHH*

AG983-961c

65	1	ATGACTTCTG	CGCCCGACTT	CAATGCAGGC	GGTACCGGTA	TCGGCAGCAA
	51	CAGCAGAGCA	ACAACAGCGA	AATCAGCAGC	AGTATCTTAC	GCCGGTATCA
	101	AGAACGAAAT	GTGCAAAGAC	AGAACGATGC	TCTGTGCCGG	TCGGGATGAC
	151	GTGCGGGTTA	CAGACAGGGA	TGCCAAATC	AATGCCCCCC	CCCCGAATCT
	201	GCATACCGGA	GACTTTCCAA	ACCCAAATGA	CGCATAACAAG	AATTTGATCA

251	ACCTCAAACC	TGCAATTGAA	GCAGGGCTATA	CAGGACGCGG	GGTAGAGGTA
301	GGTATCGTCG	ACACAGGCGA	ATCCGTCGGC	AGCATATCCT	TTCCCGAACT
351	GTATGGCAGA	AAAGAACACG	GCTATAACGA	AAATTACAAA	AACTATAACGG
401	CGTATAATGCG	GAAGGAAGCG	CCTGAAGACG	GAGGCGGTAA	AGACATITGAA
451	GCTTCTTTCG	ACGATGAGGC	CGTTATAGAG	ACTGAAGCAA	AGCCGACCGGA
501	TATCCGCCAC	GTAAAAGAAA	TCGGACACAT	CGATTGGTC	TCCCATATTA
551	TTGGCGGGCG	TTCCGTGGAC	GGCAGACCTG	CAGGCGGTAT	TGCGCCCGAT
601	CGGACGCTAC	ACATAATGAA	TACGAATGAT	GAAACCAAAGA	ACGAAATGAT
651	GGTTGCAGCC	ATCCGCAATG	CATGGGTCAA	GCTGGGCGAA	CGTGGCGTGC
701	GCATCGTCAA	TAACAGTTT	GGAAACAACAT	CGAGGGCAGG	CACTGCCGAC
751	CTTTTCCAAA	TAGCCAATT	GGAGGAGCAG	TACCGCCAAG	CGTTGCTCGA
801	CTATTCCGGC	GGTGATAAAA	CAGACGAGGG	TATCCGCTG	ATGCAACAGA
851	GGCATTACGG	CAACCTGTCC	TACACACATCC	GTAATAAAAAA	CATGCTTTTC
901	ATCTTTTCGA	CAGGCAATGA	CGCACAAAGCT	CAGCCAACA	CATATGCCCT
951	ATTGCCATT	TATGAAAAAA	ACGCTCAAAA	AGGCATTATC	ACAGTCGCAG
1001	GCGTAGACCG	CAGTGGAGAA	AAGTTCAAAAC	GGGAAATGTA	TGGAGAACCG
1051	GGTACAGAAC	CGCTTGAGTA	TGGCTCCAAC	CATTGCGGAA	TTACTGCCAT
1101	GTGGTGCCTG	TCGGCACCT	ATGAAGCAAG	CGTCCGTTTC	ACCCGTACAA
1151	ACCCGATTCA	AATTGCCGGA	ACATCCTTTT	CCGCACCCAT	CGTAACCGGC
1201	ACGGCGGCTC	TGCTGCTGA	GAAATACCCG	TGGATGAGCA	ACGACAACCT
1251	GCGTACACG	TTGCTGACGA	CGGCTCAGGA	CATCGGTGCA	GTCGGCGTGG
1301	ACAGCAAGTT	CGGCTGGGA	CTGCTGGATG	CGGGTAAGGC	CATGAACCGA
1351	CCCGCGTCTT	TTCCGTTCGG	CGACTTTTAC	GCCGATACGA	AAGGTACATC
1401	CGATATTGCC	TACTCCTCC	GTAACGACAT	TTCAGGCACG	GGCGGGCTGA
1451	TCAAAAAAGG	CGCGAGCCAA	CTGCAACTGC	ACGGCAACAA	CACCTATACG
1501	GGCAAAACCA	TTATCGAAGG	CGCTTCGCTG	GTGTTGTACG	GCAACAAACAA
1551	ATCGGATATG	CGCGTCGAAA	CCAAAGGTGC	GCTGATTAT	AACGGGGCGG
1601	CATCCGGCGG	CAGCCTGAAC	AGCGACGGCA	TTGTCTATCT	GGCAGATAAC
1651	GACCAATCCG	GCGCAAACGA	AACCGTACAC	ATCAAAGGCA	GTCTGCAGCT
1701	GGACGGCAA	GGTACGCTGT	ACACACGTTT	GGGCAAACGT	CTGAAAGTGG
1751	ACGGTACGGC	GATTATCGGC	GGCAAGCTGT	ACATGTCGGC	ACGCGGCAAG
1801	GGGGCAGGCT	ATCTCAACAG	TACCGGACGA	CGTGTCCCT	TCCTGAGTGC
1851	CGCCAAATC	GGGCAGGATT	ATTCTTTCTT	CACAAACATC	GAAACCGACG
1901	GCGGCCTGCT	GGCTTCCCCT	GACAGCGTCG	AAAAAAACAGC	GGGCAGTGAA
1951	GGCGACACGC	TGTCCTATTA	TGTCGGTCGC	GGCAATCGGG	CACGGACTGC
2001	TTCCGGCAGCG	GCACATTCG	CGCCCGCCCG	TCTGAAACAC	GCCGTAGAAC
2051	AGGGCGGCAG	CAATCTGGAA	AACCTGATGG	TCGAACTGGA	TGCTCCGAA
2101	TCATCCGCAA	CACCCGAGAC	GGTTGAAAAC	GCGGCAGCCG	ACCGCACAGA
2151	TATGCCGGGC	ATCCGCCCC	ACGGCGAAC	TTTCCGCGCA	GGGGCAGCCG
2201	TACAGCATGC	GAATGCCGCC	GACGGTGTAC	GCATCTCAA	CAGTCTCGCC
2251	GCTACCGTCT	ATGCGACAG	TACCGCCGCC	CATGCCGATA	TGCAGGGACG
2301	CCGCTCTGAA	GCGTATCGG	ACGGGTTGGA	CCACAACGGC	ACGGGTCTGC
2351	GCGTCATCGC	GCAAACCCAA	CAGGACGGTG	GAACGTGGGA	ACAGGGCGGT
2401	GTGAAAGGCA	AAATGCCGG	CAGTACCCAA	ACCGTCGGCA	TTGCGCGGAA
2451	AACCGGCAGA	AATACGACAG	CAGCCGCCAC	ACTGGGCATG	GGACGCAGCA
2501	CATGGAGCGA	AAACAGTGC	AATGCAAAAAA	CCGACAGCAT	TAGTCGTTT
2551	GCAGGCATAC	GGCACGATGC	GGGCAGTATC	GGCTATCTCA	AAGGCCTGTT
2601	CTCCTACGGA	CGCTACAAAA	ACAGCATCAG	CCGCAGCAC	GGTGGGGACG
2651	AAACATGCGGA	AGGCAGCGTC	AACGGCACGC	TGATGCGAGCT	GGGCGCACTG
2701	GGCGGTGTCA	ACGTTCCGTT	TGCCCGAACG	GGAGATTTGA	CGGTGCGAAGG
2751	CGGTCTGCGC	TACGACCTGC	TCAAACAGGA	TGCAITTCGCC	GAAAAGGCA
2801	GTGCTTTGGG	CTGGAGCGGC	AACAGCCTCA	CTGAAGGCAC	GCTGGTCGGA
2851	CTCGGGGGTC	TGAAGCTGTC	GCAACCCCTG	ACCGATAAAAG	CCGTCTGTT
2901	TGCAACGGCG	GGCGTGGAA	GCGAACCTGAA	CGGACGCGAC	TACACGGTAA
2951	CGGGCGGCTT	TACCGGGCG	ACTGCAGCAA	CCGGCAAGAC	GGGGGCACGC
3001	AATATGCCGC	ACACCCGCT	GGTTGCCGGC	CTGGGCGCGG	ATGTCGAATT
3051	CGGCAACGGC	TGGAACGGCT	TGGCACGTTA	CAGCTACGCC	GGTTCACAAAC
3101	AGTACGGCAA	CCACAGCGGA	CGAGTCGGCG	TAGGCTACCG	GTTCCCTCGAG
3151	GGTGGCGGAG	GCACTGGATC	CGCCACAAAC	GACGACGATG	TTAAAAAAAGC
3201	TGCCACTGTG	GCCATTGCTG	CTGCTACAA	CAATGGCAA	GAAATCAACG
3251	GTTCAAAGC	TGGAGAGACC	ATCTACGACA	TTGATGAAGA	CGGCACAATT
3301	ACCAAAAAAG	ACGCAACTGC	AGCCGATGTT	GAAGCCGACG	ACTTTAAAGG
3351	TCTGGGTCTG	AAAAAAAGTCG	TGACTAACCT	GACCAAAACC	GTCAATGAAA
3401	ACAAACAAAA	CGTCGATGCC	AAAGTAAAAG	CTGAGAACATC	TGAAATAGAA
3451	AAGTTAACAA	CCAAGTTAGC	AGACACTGAT	GCCGCCTT	CAGATACTGA
3501	TGCCGCTCTG	GATGCAACCA	CCAACGCCCT	GAATAAAITG	GGAGAAAATA
3551	TAACGACATT	TGCTGAAGAG	ACTAAGACAA	ATATCGTAA	AATTGATGAA

3601 AAATTAGAAG CCGTGGCTGA TACCGTCGAC AAGCATGCCG AAGCATTCAA
 3651 CGATATCGCC GATTCAATTGG ATGAAACCAA CACTAAGGCA GACGAAGCCG
 3701 TCAAAACCGC CAATGAAGCC AAACAGACGG CGAAGAAC CAAACAAAAC
 3751 GTCGATGCCA AACTAAAAGC TGCAGAAACT GCAGCAGGCA AAGCCGAAGC
 5 3801 TGCGCTGGC ACAGCTAATA CTGCAGCCGA CAAGGCCGAA GCTGTCGCTG
 3851 CAAAAGTTAC CGACATCAAA GCTGATATCG CTACGAACAA AGATAATATT
 3901 GCTAAAAAAAG CAAACAGTGC CGACGTGTAC ACCAGAGAAG AGTCTGACAG
 3951 CAAATTGTC AGAATTGATG GTCTGAACGC TACTACCGAA AAATTGGACA
 10 4001 CACGCTTGGC TTCTGCTGAA AAATCCATTG CCGATCACGA TACTCGCCTG
 4051 AACGGTTTGG ATAAAACAGT GTCAGACCTG CGCAAAGAAA CCCGCCAAGG
 4101 CCTTGCAGAA CAAGCCGCGC TCTCCGGTCT GTTCCAACCT TACAACGTGG
 4151 GTCTCGAGCA CCACCACAC CACCACTGA

15 1 MTSAPDFNAG GTGIGSNSRA TTAKSAAVSY AGIKNEMCKD RSMILCAGRDD
 51 VAVTDRDAKI NAPPPNLHTG DFPNPNDAYK NLINLKPAIE AGYTGRGVEV
 101 GIVDTGESVG SISFPELYGR KEHGYNENYK NYTAYMRKEA PEDGGKDIE
 151 ASFDDEAVIE TEAKPTDIRH VKEIGHIDLV SHIIGGRSVD GRPAGGTAPD
 201 ATLHIMNTND ETKNEMMVAI IRNAWVKLGE RGVRIVNNSF GTTSRAGTAD
 251 LFQIANSEEQ YRQALLDYSG GDKTDEGIRL MQQSDYGNLIS YHIRNKNMLF
 301 IFSTGNDQAQ AQPNTYALLPF YEKDAQKGII TVAGVDRSGE KFKREMYGEP
 351 GTEPLEYGSN HCGITAMWCL SAPYEASVRF TRTNPIQIAG TSFSAPIVTG
 401 TAALLLQKYP WMSNDNLRRTT LLTTAQDIGA VGVDSKFGWG LLDAGKAMNG
 451 PASFPFGDFT ADTKGTSIDIA YSFNRDISGT GGLIKKGGSQ LQLHGNNTYT
 501 GKTIEGGSL VLYGNNKSDM RVETKGALIY NGAASGGSLN SDGIVYLA
 551 DQSGANETVH IKGSLQLDGK GTLYTRLGKL LKVDGTAIIIG GKLYMSARGK
 601 GAGYLNSTGR RVPFLSAKI GQDYSFFTNI ETDGGLLASL DSVEKTAGSE
 651 GDTLSYYVRR GNAARTASAA AHSAPAGLKH AVEQGGSNLE NLMVELDASE
 701 SSATPETVET AAADRTDMPG IRPYGATFRA AAAVQHANAA DGVRIFNSLA
 751 ATVYADSTAA HADMQGRRLK AVSDGLDHNG TGLRVIAQTQ QDGTTWEQGG
 30 801 VEGKMRGSTQ TVGIAAKTGE NTTAAATLGM GRSTWSENSA NAKTDSISLF
 851 AGIRHDAGDI GYLKGLFSYG RYKNSISRST GADEHAEGSV NGTLMQLGAL
 901 GGVNPFAAT GDLTVEGGLR YDLLKQDAFA EKGSALGWSG NSLTEGTLVG
 951 LAGLKLSQLP LDKAVLFBATA GVERDLNGRD YTWTGGFTGA TAATGKTGAR

35 1001 NMPHTRLVAG LGADVEFGNG WNGLARYSYA GSKQYGNHSG RVGVGYRFLE
 1051 GGGGTGSATN DDDVKKAAATV AIAAAYNNGQ EINGFKAGET IYDIDEDGTI
 1101 TKKDATAADV EADDFKGLGL KKVVTNLTKT VNEENKQNVDV KVKAEESEIE
 1151 KLTTKLADTD AALADTDAAL DATTNALNKL GENITTFAAEE TKTNIVKIDE
 1201 KLEAVADTVL KHAEEAFNDIA DSLDETNTKA DEAVKTANEAA KQTAEEETKQN
 40 1251 VDAKVKAAT AAGKAEAAAG TANTAADKAE AVAAKVTDIK ADIATMNDNI
 1301 AKKANSADVY TREESDSKFV RIDGLNATTE KLDTRLASAE KSIADHDTRL
 1351 NGLDKTVSDL RKETRQGLAE QAALSGLFQP YNVGLEHHHH HH*

ΔG741 and hybrids

Bactericidal titres generated in response to ΔG741 (His-fusion) were measured against various strains, including the homologous 2996 strain:

	2996	MC58	NGH38	F6124	BZ133
ΔG741	512	131072	>2048	16384	>2048

45 As can be seen, the ΔG741-induced anti-bactericidal titre is particularly high against heterologous strain MC58.

ΔG741 was also fused directly in-frame upstream of proteins 961, 961c, 983 and ORF46.1:

ΔG741-961
 1 ATGGTCGCCG CCGACATCGG TGCAGGGCTT GCCGATGCAC TAACCGCACC
 51 GCTCGACCAT AAAGACAAAG GTTTCAGTC TTTGACGCTG GATCAGTCCG
 101 TCAGGAAAAA CGAGAAACTG AAGCTGGCGG CACAAGGTGC GGAAAAAAACT
 151 TATGGAAACG GTGACAGCCT CAATACGGGC AAATTGAAGA ACGACAAGGT
 201 CAGCCGTTTC GACTTTATCC GCCAAATCGA AGTGGACGGG CAGCTCATTA

251 CCTTGGAGAG TGGAGAGTTC CAAGTATACA AACAAAGCCA TTCCGCCTTA
 301 ACCGCCTTTC AGACCGAGCA AATACAAGAT TCGGAGCATT CCGGAAAGAT
 351 GTTTCGAA CGCCAGTCA GAATCGCGA CATAGCGGGC GAACATACAT
 401 CTTTGACAA GCITCCGAA GCGCGCAGGG CGACATATCG CGGACCGCG
 451 TTCGGTTCAG ACGATGCCGG CGGAAAATCG ACCTACACCA TAGATTCGC
 501 CGCCAAGCAG GGAAACGGCA AAATCGAAC TTTGAAATCG CCAGAACTCA
 551 ATGTCGACCT GGCGCGGCC GATATCAAGC CGGATGGAAA ACGCCATGCC
 601 GTCATCAGCG GTTCCGTCT TTACAACCAA GCGGAGAAA GCAGTTACTC
 651 CCTCGGTATC TTTGGCGAA AAGCCCAGGA AGTTGCCGGC AGCGCGGAAG
 701 TGAAAACCGT AAACGGCATA CGCCATATCG GCCTTGCGGC CAAGCAACTC
 751 GAGGGTGGCG GAGGCAGTGG ATCCGCCACA AACGACGACG ATGTTAAAAA
 801 AGCTGCCACT GTGGCCATTG CTGCTGCCA CAACAATGGC CAAGAAATCA
 851 ACGGTTTCAA AGCTGGAGAG ACCATCTACG ACATTGATGA AGACGCCACA
 901 ATTACCAAAA AAGACGCAAC TGCAGCCGAT GTGAAAGCCG ACGACTTTAA
 951 AGGTCTGGGT CTGAAAAAAAG TCCTGACTAA CCTGACCAAA ACCGTCATG
 1001 AAAACAAACA AAACGTCGAT GCCAAAGTAA AAGCTGCAGA ATCTGAAATA
 1051 GAAAAGTTAA CAACCAAGTT AGCAGACACT GATGCCGCTT TAGCAGATAC
 1101 TGATGCCGCT CTGGATGCAA CCACCAACGC CTTGAATAAA TTGGGAGAAA
 1151 ATATAACGAC ATTTGCTGAA GAGACTAAGA CAAATATCGT AAAAATTGAT
 1201 GAAAATTAG AAGCGTGGC TGATACCGTC GACAAGCATG CCGAAGCATT
 1251 CAACGATATC GCGGATTCTAT TGGATGAAAC CAACACTAAAG GCAGACGAAG
 1301 CCGTAAACAC CGCCAATGAA GCCAACACAGA CGGCCGAAGA AACCAAACAA
 1351 AACGTCGATG CCAAAGTAAAG AGCTGCAGAA ACTGCAAGCAG GCAAAAGCCGA
 1401 AGCTGCCGCT GGCACAGCTA ATACTGCAGC CGACAAGGCC GAAGCTGTCG
 1451 CTGCAAAAGT TACCGACATC AAAGCTGATA TCGCTACGAA CAAAGATAAT
 1501 ATTGCTAAAAA AAGCAACAG TGCCGACGTG TACACCAAGA AAGAGTCTGA
 1551 CAGCAAATTG GTCAGAAATTG ATGGTCTGAA CGCTACTACC GAAAATTGG
 1601 ACACACGCTT GGCTTCTGCT GAAAATCCA TTGCCGATCA CGATACTCGC
 1651 CTGAACGGTT TGGATAAAAC AGTGTCAAGAC CTGCGCAAAG AAACCCGCCA
 1701 AGGCCTTGCA GAACAAGCCG CGCTCTCCGG TCTGTTCAA CCTTACAACG
 1751 TGGGTGGTT CAATGTAACG GCTGCAGTCG GCGGCTACAA ATCCGAATCG
 1801 GCAGTCGCCA TCGGTACCGG CTTCCGCTTT ACCGAAAACCT TTGCCGCCAA
 1851 AGCAGGCGTG GCAGTCGGCA CTTCGTCCGG TTCTTCCGCA GCCTACCATG
 1901 TCGCGTCAA TTACGAGTGG CTCGAGCACC ACCACCAACCA CCACTGA
 35
 1 MVAADIGAGL ADALTAPLDH KDKGLQSLTL DQSVRKNEKL KLAQGAEKT
 51 YNGDSDLNTG KLNKNDKVSRF DFIRQIEVDG QLITTLESGEF QVYKQSHSAL
 101 TAFQTEQIJD SEHSGKMKVAK RQFRIGDIAG EHTSFDKLPE GGRATYRGTA
 151 FGSDDAGGKL TYTIDFAAKQ GNGKIEHLKS PELNVDLAAA DIKPDGKRHA
 201 VISGSVLYNQ AEKGSYSLGI FGGKAQEVAQ SAEVKTVNGI RHIGLAAKQL
 251 EGGGGTGSAT NDDDVKKAAAT VAIAAAYNNG QBLINGFKAGE TIYDIDEDGT
 301 ITKKKDATAAD VEADDFKGLG LKKVVTNLTK TVNENKQNVD AKVKAEESEI
 351 EKLTTKLADT DAALADTDAA LDATTNALNK LGENITTFAE ETKTNIVKID
 401 EKLEAVADTV DKHAEAFNDI ADSLDETNTK ADEAVAKTANE AKQTAETKQ
 451 NVDAKVKAEE TAAGKAEAAA GTANTAADKA EAVAATVTDI KADIATNKDN
 501 IAKKANSADV YTREESDSKF VRIDLGNATT EKLDTRLASA EKSIADHDTR
 551 LNGLDKTVSD LRKETRQGLA EQAALSGLFQ PYNVGRFNVT AAVGGYKSES
 601 AVAIGTGFRF TENFAAKAGV AVGTSGGSSA AYHVGVNUEW LEHHHHHH*
 50
ΔG741-961c
 1 ATGGTCGCCG CCGACATCGG TGCAGGGCTT GCCGATGCAC TAACCGCACC
 51 GCTCGACCAT AAAGACAAAG GTTTCGAGTC TTTGACGCTG GATCAGTCCG
 101 TCAGGAAAAA CGAGAAAATG AAGCTGGCGG CACAAGGTGC GGAAAAAAACT
 151 TATGGAAACG GTGACAGCCT CAATACGGGC AAATTGAAAGA ACGACAAGGT
 201 CAGCGTTTC GACTTTATCC GCCAAATCGA AGTGGACGGG CAGCTCATTA
 251 CCTTGGAGAG TGGAGAGTTC CAAGTATACA AACAAAGCCA TTCCGCCTTA
 301 ACCGCCTTTC AGACCGAGCA AATACAAGAT TCGGAGCATT CCGGAAAGAT
 351 GTTTGCGAAA CGCCAGTCA GAATCGCGA CATAGCGGGC GAACATACAT
 401 CTTTGACAA GCTTCCGAA GGCAGCAGGG CGACATATCG CGGACCGCG
 451 TTTCGGTTCAG ACGATGCCGG CGGAAAATCG ACCTACACCA TAGATTCGC
 501 CGCCAAGCAG GGAAACGGCA AAATCGAAC TTTGAAATCG CCAGAACTCA
 551 ATGTCGACCT GGCGCGGCC GATATCAAGC CGGATGGAAA ACGCCATGCC
 601 GTCATCAGCG GTTCCGTCT TTACAACCAA GCGGAGAAA GCAGTTACTC
 651 CCTCGGTATC TTTGGCGAA AAGCCCAGGA AGTTGCCGGC AGCGCGGAAG
 701 TGAAAACCGT AAACGGCATA CGCCATATCG GCCTTGCGGC CAAGCAACTC
 751 GAGGGTGGCG GAGGCAGTGG ATCCGCCACA AACGACGACG ATGTTAAAAA

5	801	AGCTGCCACT	GTGGCCATTG	CTGCTGCCTA	CAACAATGGC	CAAGAAATCA	
	851	ACGGTTCAA	AGCTGGAGAG	ACCATCTACG	ACATTGATGA	AGACGGCACA	
	901	ATTACCAAAA	AAGACGCAAC	TGCAGCCGAT	GTGAAGCCG	ACGACTTTAA	
	951	AGGTCTGGGT	CTGAAAAAAG	TCGTAAGTAA	CCTGACCAAA	ACCGTCAATG	
10	1001	AAAACAAACA	AAACGTCGAT	GCCAAAGTAA	AAGCTGCAGA	ATCTGAAATA	
	1051	GAAAAGTTAA	CAACCAAGTT	AGCAGACACT	GATGCCGCTT	TAGCAGATAC	
	1101	TGATGCCGCT	CTGGATGCAA	CCACCAACGC	CTTGAATAAA	TTGGGAGAAA	
	1151	ATATAACGAC	ATTTGCTGAA	GAGACTAAGA	CAAATATCGT	AAAAATTGAT	
15	1201	GAAAAATTAG	AAGCGTGGC	TGATACCGTC	GACAAGCATG	CCGAAGCATT	
	1251	CAACGATATC	GCCGATTCCAT	TGGATGAAAC	CAACACTAAG	GCAGACGAAG	
	1301	CCGTCAAAAC	CGCCAATGAA	GCCAAACAGA	CGGCCGAAGA	AACCAAACAA	
	1351	AACGTGATG	CCAAAGTAA	AGCTGCAGAA	ACTGCAGCAG	GCAAAGCCGA	
	1401	AGCTGCCGCT	GGCACAGCTA	ATACTGCAGC	CGACAAGGCC	GAAGCTGTCG	
	1451	CTGCAAAGT	TACCGACATC	AAAGCTGATA	TCGCTACGAA	CAAAGATAAT	
20	1501	ATTGCTAAAA	AAGCAAACAG	TGCCGACGTG	TACACCAGAG	AAGAGTCTGA	
	1551	CAGCAAATT	GTCAGAAATTG	ATGGTCTGAA	CGCTACTACC	AAAAATTGG	
	1601	ACACACGCTT	GGCCTCTGCT	AAAAATCCA	TTGCCGATCA	CGATACTCGC	
	1651	CTGAACGGTT	TGGATAAAAC	AGTGTCAAGAC	CTGCGCAAAG	AAACCGCCA	
	1701	AGGCCTTGCA	GAACAAGCCG	CGCTCTCCGG	TCTGTTCCAA	CCTTACAACG	
	1751	TGGGTCTCGA	GCACCAAC	CACCAACACT	GA		
25	1	MVAADIGAGL	ADALTAPLDH	KDKGLQSLTL	DQSVRKNEKL	KLAAQGAEKT	
	51	YNGNDSLNTG	KLKNDKVSRF	DFIRQIEVDG	QLITLESGEF	QVYKQSHSAL	
	101	TAFQTEQIJD	SEHSGKMKVAK	RQFRIGDIAG	EHTSFDKLPE	GRATYRGT	
	151	FGSDDAGGKL	TYTIDFAAQ	GNGKIEHLKS	PELNVDLAAA	DIKPDKRHA	
	201	VISGSVLYNQ	AEKGSYSLGI	FFGKAQEVAG	SAEVKTVNGI	RHIGLAAKQL	
	251	EGGGGGTGSAT	NDDDVKKAA	VAIAAAYNNG	QEINGFKAGE	TIYDIDEDGT	
	301	ITKKDATAAD	VEADDFKGLG	LKKVVTNLTK	TVNENKQNV	AKVAAESEI	
30	351	EKLTTKLADT	DAALADTAA	LDATTNALNK	LGENITTFAE	ETKTNIVKID	
	401	EKLEAVADTV	DKHAEAFNDI	ADSLDETNK	ADEAVKTANE	AKQTAETKQ	
	451	NVDAKVAAE	TAAGKAEAAA	GTANTAADKA	EAVAAKVTDI	KADIATNKN	
	501	IAKKANSADV	YTREESDSKF	VRIDGLNATT	EKLDTRLASA	EKSIADHDTR	
	551	LNGLDKTVSD	LRKETRQGLA	EQAALSGLFQ	PYNVGLEHHH	HHH*	
35	<u>AG741-983</u>						
40	1	ATGGTCGCCG	CCGACATCGG	TGCGGGGCTT	GCCGATGCAC	TAACCGCACC	
	51	GCTCGACCAT	AAAGACAAAG	GTTTGAGTC	TTTGACGCTG	GATCAGTCG	
	101	TCAGGAAAAA	CGAGAAACTG	AAGCTGGCGG	CACAAGGTGC	GGAAAAAAACT	
	151	TATGGAAACG	GTGACAGCCT	CAATACGGGC	AAATTGAAGA	ACGACAAGGT	
	201	CAGCCGTTTC	GACTTTATCC	GCCAAATCGA	AGTGGACGGG	CAGCTCATTA	
	251	CCTTGGAGAG	TGGAGAGTTC	CAAGTATACA	ACAAAGGCCA	TTCCGCCCTTA	
	301	ACCGCCTTTC	AGACCGAGCA	AATACAAGAT	TCGGAGCATT	CCGGGAAGAT	
	351	GGTTGCGAAA	CGCCAGTCA	GAATCGGCCA	CATAGCGGGC	GAACATACAT	
45	401	CTTTTGACAA	GCTTCCCAGA	GGCGGCAGGG	CGACATATCG	CGGGACGGCG	
	451	TTCCGGTTCA	ACGATGCCG	CGGAAAATG	ACCTACACCA	TAGATTTCGC	
	501	CGCCAAGCAG	GGAAACGGCA	AAATCGAACAA	TTTGAAATCG	CCAGAACTCA	
	551	ATGTCGACCT	GGCCGCCGCC	GATATCAAGC	CGGATGGAAA	ACGCCATGCC	
50	601	GTCATCAGCG	GTTCGTCCT	TTACAACCAA	GCCGAGAAAG	GCAGTTACTC	
	651	CCTCGGTATC	TTTGGCGAA	AAGCCCAGGA	AGTTGCCGGC	AGCGCGGAAG	
	701	TGAAAACCGT	AAACGGCATA	CGCCATATCG	GCCTTGCCTCG	CAAGCAACTC	
	751	GAGGGATCCG	GCGGAGGGCG	CACTTCTGCG	CCCGACTTC	ATGCAGGCC	
	801	TACCGGTATC	GGCAGCAAC	GCAGAGCAAC	ACAGCGAAA	TCAGCAGCAG	
	851	TATCTTACGC	CGGTATCAAG	AACGAAATGT	GCAAAAGACAG	AAGCATGCTC	
55	901	TGTGCCGGTC	GGGATGACGT	TGCGGTTACA	GACAGGGATG	CCAAAATCAA	
	951	TGCCCCCCCC	CCGAATCTGC	ATACCGGAGA	CTTTC	CCAAATGACG	
	1001	CATACAAGAA	TTTGATCAAC	CTCAAACCTG	CAATTGAAAGC	AGGCTATACA	
	1051	GGACGCCGGG	TAGAGGTAGG	TATCGTCAC	ACAGGGCAAT	CCGTCCGCAG	
60	1101	CATATCCTT	CCCGAACTGT	ATGGCAGAAA	AGAACACGGC	TATAACGAAA	
	1151	ATTACCAAAA	CTATACGGC	TATATGCCGA	AGGAAGCGCC	TGAAGACGGA	
	1201	GGCGGTAAAG	ACATTGAAGC	TTCTTCGAC	GATGAGGCCG	TTATAGAGAC	
	1251	TGAAGCAAAAG	CCGACGGATA	TCCGCCACGT	AAAAGAAATC	GGACACATCG	
	1301	ATTGGGTCTC	CCATATTATT	GGCGGGCGTT	CCGTGGACGG	CAGACCTGCA	
	1351	GGCGGTATTG	CGCCCGATGC	GACGCTACAC	ATAATGAATA	CGAATGATGA	
65	1401	AACCAAGAAC	GAAATGATGG	TTGCAGCCAT	CCGCAATGCA	TGGGTCAAGC	
	1451	TGGCGAACG	TGGCGTGC	ATCGTCAATA	ACAGTTTTGG	AACAAACATCG	
	1501	AGGGCAGGCA	CTGCCGACCT	TTTCCAATA	GCCAATTCCG	AGGAGCAGTA	

1551	CCGCAAGCG	TTGCTCGACT	ATTCCGGCGG	TGATAAAAACA	GACGAGGGTA	
1601	TCCGCCTGAT	GCAACAGAGC	GATTACGGCA	ACCTGTCTA	CCACATCCGT	
1651	AATAAAAACA	TGCTTTTCAT	CTTTTCGACA	GGCAATGACG	CACAAGCTCA	
1701	GCCCAACACA	TATGCCCTAT	TGCCATTTTA	TGAAAAAAGAC	GCTCAAAAAG	
1751	GCATTATCAC	AGTCGCAGGC	GTAGACCGCA	GTGGAGAAAAA	GTTCAAACGG	
1801	GAAATGTATG	GAGAACCGGG	TACAGAACCG	CTTGAGTATG	GCTCCAACCA	
1851	TTGCGGAATT	ACTGCCATGT	GGTGCCTGTC	GGCACCTAT	GAAGCAAGCG	
1901	TCCGTTTCAC	CCGTACAAAC	CCGATTCAAA	TTGCCGGAAC	ATCCTTTCC	
1951	GCACCCATCG	TAACCGGCAC	GGCGGCTCTG	CTGCTGCAGA	AATACCCGTG	
2001	GATGAGCAAC	GACAACCTGC	GTACCACGTT	GCTGACGACG	GCTCAGGACA	
2051	TCGGTGCAGT	CGCGTGGAC	AGCAAGTCG	GCTGGGGACT	GCTGGATGCG	
2101	GGTAAGGCCA	TGAACGGACC	CGCGTCCTT	CCGTTCGCG	ACTTTACCGC	
2151	CGATACGAAA	GGTACATCCG	ATATTGCCTA	CTCCTTCCGT	AACGACATT	
2201	CAGGCACGGG	CGGCCTGATC	AAAAAAGGCG	GCAGCCAACT	GCAACTGCAC	
2251	GGCAACACA	CCTATACGGG	CAAACATT	ATCGAAGGCG	GTTCGCTGGT	
2301	GTTGTACGGC	AACAACAAAT	CGGATATGCG	CGTCGAAACC	AAAGGTGCGC	
2351	TGATTTATAA	CGGGCGGCA	TCCGGCGGCA	GCCTGACAG	CCACGGCATT	
2401	GTCTATCTGG	CAGATACCGA	CCAATCCGGC	GCAAAACGAAA	CCGTACACAT	
2451	CAAAGGCAGT	CTGCAGCTGG	ACGGCAAAAG	TACGCTGTAC	ACACGTTGG	
2501	GCAAACGTCT	GAAAGTGGAC	GGTACGGCGA	TTATCGGCGG	CAAGCTGTAC	
2551	ATGTCGGCAC	GCGGCAAGGG	GGCAGGCTAT	CTCAACAGTA	CCGGACGACG	
2601	TGTTCCCTTC	CTGAGTGGCG	CCAAAATCGG	GCAGGATTAT	TCTTTCTTCA	
2651	CAAACATCGA	AACCGACGGC	GGCCTGCTGG	CTTCCCTCGA	CAGCGTCGAA	
2701	AAAACAGCGG	GCAGTGAAGG	CGCACACGCTG	TCCTATTATG	TCCGTCGCGG	
2751	CAATGCGGCA	CGGACTGCTT	CGGCAGCGGC	ACATTCCGCG	CCCGCCGGTC	
2801	TGAAACACGC	CGTAGAACAG	GGCGGCAGCA	ATCTGAAA	CCTGATGGTC	
2851	GAACTGGATG	CCTCCGAATC	ATCCGCAACA	CCCGAGACGG	TTGAAACTGC	
2901	GGCAGCCGAC	CGCACAGATA	TGCCGGGCAT	CCGCCCCCTAC	GGCGCAACCT	
2951	TCCGCGCAGC	GGCAGCCGTA	CAGCATGCGA	ATGCCGCGCA	CGGTGTACGC	
3001	ATCTTCAACA	GTCTCGCCGC	TACCGTCTAT	GCGACAGTA	CCGCGGCCA	
3051	TGCCGATATG	CAGGGACGCC	GCCTGAAAGC	CGTATCGGAC	GGGTTGGACC	
3101	ACAACGGCAC	GGGTCTGCGC	GTCATCGCGC	AAACCCAACA	GGACGGTGG	
3151	ACGTGGGAATC	AGGGCGGTGT	TGAAGGCAAA	ATGCGCGGC	GTACCCAAAC	
3201	CGTGGCATT	CGCCGCAAAA	CCGGCGAAAA	TACGACAGCA	GCCGCCACAC	
3251	TGGGCATGGG	ACGCAGCAC	TGGAGCGAAA	ACAGTGCAAA	TGCAAAAC	
3301	GACAGCATT	GTCTGTTGC	AGGCATACGG	CACGATGCGG	GCGATATCGG	
3351	CTATCTAAA	GGCCTGTCT	CCTACGGACG	CTACAAAAAC	ACCATCAGCC	
3401	GCAGCACCGG	TGCGGACGAA	CATGCGGAAG	GCAGCGTCAA	CCGCACGCTG	
3451	ATGCAGCTGG	GCGCACTGGG	CGGTGTCAAC	GTTCGTTTG	CCGCAACGGG	
3501	AGATTGACG	GTCGAAGGCG	GTCGCGCTA	CGACCTGCTC	AAACAGGATG	
3551	CATTGCGCGA	AAAAGGCAGT	GCTTGGGCT	GGAGCGGCAA	CAGCCCTCACT	
3601	GAAGGCACGC	TGGTCGGACT	CGCGGGTCTG	AAGCTGTCCG	AACCCCTTGAG	
3651	CGATAAACGC	GTCCTGTTG	CAACGGCGGG	CGTGGAAACCG	GACCTGAACG	
3701	GACCGCGACTA	CACGGTAACG	GGCGGCTTTA	CCGGCGCGAC	TGCAGCAACC	
3751	GGCAAGACGG	GGGCACGCAA	TATGCCGCAC	ACCCGCTCTGG	TTGCCGGCCT	
3801	GGGCGCGGAT	GTCGAATCG	GCAACGGCTG	GAACGGCTTG	GCACGTTACA	
3851	GCTACGCCGG	TTCCAAACAG	TACGGCAACC	ACAGCGGACG	AGTCGGCGTA	
3901	GGCTACCGGT	TCCTCGAGCA	CCACCAACCAC	CACCACTGA		
50	1	MVAADIGAGL	ADALTAPLDH	KDKGLQSLTL	DQSVRKNEKL	KLAAGQAEKT
	51	YNGDSDLNTG	KLKNNDKVSRF	DFIRQIEVDG	QLITTLESGEF	QVYKQSHSAL
	101	TAFQTEQIJD	SEHSGKVMVK	RQFRIGDIAG	EHTSFDKLPE	GGRATYRGTA
	151	FGSDDAGGKL	TYTIDFAKO	GNKGIEHLKS	PELNVDLAAA	DIKPDKRHA
	201	VISGSVLYNQ	AEKGSYSLGI	FGGKAQEVAG	SAEVKTVNGI	RHIGLAAKQL
	251	EGSGGGGTSA	PDFNAGGTGI	GSNSRATTAK	SAAVSYAGIK	NEMCKDRSML
	301	CAGRDDVAVT	DRDAKINAPP	PNLHTGDFPN	PNDAYKNLIN	LKPAIEAGYT
	351	GRGVEVGIVD	TGESVGSISF	PELYGRKEHG	YNENYKNYTA	YMRKEAPEPDG
	401	GGKDIEASFD	DEAVIETEAK	PTDIRHVKEI	GHIDLVSHII	GGRSVDGRPA
	451	GGIAPDATALH	IMNTNDETKN	EMMVAAIRNA	WVKLGERGVR	IVNNNSFTTS
	501	RAGTADLFQI	ANSEEQYRQA	LLDYSGGDKT	DEGIRLMLMQQS	DYGNLNSYHIR
	551	NKNMLFIFST	GNDAQAQPNT	YALLPFYEKD	AQKGIITVAG	VDRSGEKFKR
	601	EMYGEPTGTEP	LEYGSNHCIG	TAMWCLSAPY	EASVRFTRTN	PIQIAGTSFS
	651	APIVTGTAAL	LLQKYPWMSN	DNLRTTLLTT	AQDIGAVGVD	SKFGWGLLDA
	701	GKAMNGPASF	PFGDFTADTK	GTSDIAYSFR	NDISGTGGLI	KKGGSQLQLH
	751	GNNTYTGKTI	IEGGSLVLYG	NNKSDMRVET	KGALIYNGAA	SGGSLNSDGI
	801	VYLADTDQSG	ANETVHIKGS	LQLDGKGTL	TRLGKLLKVD	GTAIIGGKLY
	851	MSARGKGAGY	LNSTGRRVPF	LSAAKIGQDY	SFFTNIETDG	GLLASLDSVE

901 KTAGSEGDTL SYYVRRGNA RTASAAAHS PAGLKHAVEQ GGSNLENLMV
 951 ELDASESSAT PETVETAAAD RTDMPGIRPY GATFRAAAAV QHANAADGVR
 1001 IFNSLAATVY ADSTAHAADM QGRRILKAVSD GLDHNGTGLR VIAQTQODGG
 1051 TWEQGGVEGK MRGSTQTVGI AAKTIGENTTA ATALGMGRST WSENSANAKT
 5 1101 DSISLFLAGIR HDAGDIGYLK GLFSYGRYKN SISRSTGADE HAEGSVNGTL
 1151 MQLGALGGVN VPFAATGDLT VEGGLRYDLL KQDFAAEKGGS ALGWSGNSLT
 1201 EGTLVGLAGL KLSQPLSDKA VLFATAGVER DLNGRDYTVT GGFTGATAAT
 1251 GKTGARNMPH TRIVAGLGAD VEFGNGWNGL ARYSYAGSKQ YGNHSGRVGV
 10 1301 GYRFLEHHHH HH*

ΔG741-ORF46.1
 1 ATGGTCGCCG CCGACATCGG TGCGGGCTT GCCGATGCAC TAACCGCAC
 51 GCTCGACCAT AAAGACAAAG GTTTCAGTC TTTGACGCTG GATCAGTC
 101 TCAGGAAAAA CGAGAAACTG AAGCTGGCGG CACAAGGTGC GGAAAAAAACT
 15 151 TATGGAAACG GTGACAGCCT CAATACGGGC AAATTGAAGA ACGACAAAGGT
 201 CAGCGGTTTC GACTTTATC GCCAAATCGA AGTGGACGGG CAGCTCATTA
 251 CCTTGAGAG TGGAGAGTTT CAAGTATACA ACAAAAGCCA TTCCGCCCTTA
 301 ACCGCCTTTTC AGACCGAGCA AATACAAGAT TCGGAGCATT CCGGGAAAGAT
 351 GGTTGCGAAA CGCCAGTCA GAATCGGCAG CATAGCGGGC GAACATACAT
 20 401 CTTTGACAA GCTTCCCCAA GGCAGCAGGG CGACATATCG CGGGACGGCG
 451 TTCGGTTTCAG ACGATGCCCG CGGAAAACTG ACCTACACCA TAGATTTC
 501 CGCCAAGCGAG GGAAACCGCA AAATCGAACA TTTGAAATCG CCAGAACTCA
 551 ATGTCGACCT GGCGCGGCC GATATCAAGC CGGATGGAAA ACGCCATGCC
 601 GTCATCAGCG GTTCCGTCCT TTACAACCAA GCGGAGAAAG GCAGTTACTC
 25 651 CCTCGGTATC TTGGCGCAA AAGCCAGGA AGTTGCGGGC AGCGCGGAAG
 701 TGAAACCGT AAACGGCATA CGCCATATCG GCCTTGCCTC CAAGCAACTC
 751 GACGGTGGCG GAGGCAGTGG ATCCTCAGAT TTGGCAAACG ATTCTTTTAT
 801 CCGGCAGGTT CTCGACCGTC AGCATTTCGA ACCCGACGGG AAATACCACC
 851 TATTCGGCAG CAGGGGGAA CTTGGCGAGC GCAGCGGCCA TATCGGATTG
 30 901 GGAAAAATAC AAAGCCATCA GTTGGGCAAC CTGATGATTG AACAGGGCGC
 951 CATTAAAGGA AATATCGGCT ACATTTGTCG CTTTTCCGAT CACGGGCACG
 1001 AAGTCGATTC CCCCCTCGAC AACCATGCCT CACATTCCGA TTCTGATGAA
 1051 GCGGTAGTC CCGTTGACGG ATTAGCCTT TACCGCATCC ATTGGGACGG
 1101 ATACGAACAC CATCCCCCGC ACGGCTATGA CGGGCCACAG GGCGGCGGCT
 35 1151 ATCCCGCTCC CAAAGGGCGG AGGGATATAT ACAGCTACGA CATAAAAGGC
 1201 GTTGCCAAA ATATCCGCT CAACCTGACC GACAACCGCA GCACCGGACA
 1251 ACGGTTGCC GACCGTTCC ACAATGCCGG TAGTATGCTG ACGCAAGGAG
 1301 TAGGCACGG ATTCAAACGC GCCACCCGAT ACAGCCCCGA GCTGGACAGA
 1351 TCGGGCAATG CCGCCGAAGC CTTCAACGGC ACTGCAGATA TCGTTAAAAA
 40 1401 CATCATCGC GCGGCAGGAG AAATTGTCGG CGCAGGGCAT GCGGTGCAGG
 1451 GCATAAGCGA AGGCTCAAC ATTGCTGTCA TGCACGGCTT GGGTCTGCTT
 1501 TCCACCGAAA ACAAGATGGC GCGCATCAAC GATTGGCAG ATATGGCGCA
 1551 ACTCAAAGAC TATGCCGAG CAGCCATCCG CGATGGGCA GTCCAAACACC
 1601 CCAATGCCGC ACAAGGCATA AAAGCCGTCA GCAATATCTT TATGGCAGCC
 45 1651 ATCCCCATCA AAGGGATTGG AGCTGTTCGG GGAAAATACG GCTTGGCGG
 1701 CATCACGGCA CATCCTATCA AGCGGTGCGA GATGGGCGCG ATCGCATTGC
 1751 CGAAAAGGGAA ATCCGCCGTC AGCGACAATT TTGCCGATGC GGCATACGCC
 1801 AAATAACCGT CCCCCTACCA TTCCCGAAAT ATCCGTTCAA ACTTGGAGCA
 1851 GCGTTACGGC AAAGAAAACA TCACCTCCTC AACCGTGCCTG CCGTCAAACG
 50 1901 GCAAAAATGT CAAACTGGCA GACCAACGCC ACCCGAAGAC AGGCGTACCG
 1951 TTTGACGGTA AAGGGTTCC GAATTTCGAG AACGACGTGA AATATGATAC
 2001 GCTCGAGCAC CACCAACCACC ACCACTGA

 1 MVAADIGAGL ADALTAFLDH KDKGLQSLTL DQSVRKNEKL KLAAGQAETK
 51 YGNGLDSLNTG KLNKNDKVSF DFIRQIEVDG QLITLESGEF QVYKQSHSAL
 101 TAFQTEQIJD SEHSGKMKVAK RQFRIGDIAG EHTSFDKLPE GGRATYRGTA
 151 FGSDDAGGKL TYTIDFAAKQ GNGKIEHLKS PELNVDLAAA DIKPDGKRHA
 201 VISGSVLYNQ AEKGSYSLGI FGGKAQEVAQ SAEVKTIVNGI RHIGLAAKQL
 251 DGGGGTGSSD LANDSFIRQV LDRQHFEPDG KYHILFGSRGE LAERSGHIGL
 60 301 GKIQSHQLGN LMIQQAAIKG NIGYIVRFSD HGHEVHSPPD NHASHSDSDE
 351 AGSPVDFGFL YRIHWDGYEH HPADGYDGPQ GGGYPAPKGA RDIYSYDIKG
 401 VAQNIRLNLT DNRSTGQLA DRFHAGMSML TQGVGDGFKA ATRYSPEDLR
 451 SGNAEAEAFNG TADIVKNIIG AAGEIVGAGD AVQGISEGSN IAVMHGLGLL
 501 STENKMARIN DLADMAQLD YAAAIRDWA VQNPNAAQGI EAVSNIFMAA
 551 IPIKGIGAVR GKYGGLGGITA HPIKRSQMGA IALPKGKSAV SDNFADAAYA
 601 KYPSPYHSRN IRSNLEQRYG KENITSSTVP PSNGKNVKLA DQRHPKTGVP
 651 FDGKGFPNFE KHVKYDTLEH HHHHH*

Example 16 – C-terminal fusions ('hybrids') with 287/ΔG287

According to the invention, hybrids of two proteins A & B may be either NH₂-A-B-COOH or NH₂-B-A-COOH. The effect of this difference was investigated using protein 287 either C-terminal (in '287-His' form) or N-terminal (in ΔG287 form – sequences shown above) to

5 919, 953 and ORF46.1. A panel of strains was used, including homologous strain 2996. FCA was used as adjuvant:

	287 & 919		287 & 953		287 & ORF46.1	
Strain	ΔG287-919	919-287	ΔG287-953	953-287	ΔG287-46.1	46.1-287
2996	128000	16000	65536	8192	16384	8192
BZ232	256	128	128	<4	<4	<4
1000	2048	<4	<4	<4	<4	<4
MC58	8192	1024	16384	1024	512	128
NGH38	32000	2048	>2048	4096	16384	4096
394/98	4096	32	256	128	128	16
MenA (F6124)	32000	2048	>2048	32	8192	1024
MenC (BZ133)	64000	>8192	>8192	<16	8192	2048

Better bactericidal titres are generally seen with 287 at the N-terminus (in the ΔG form)

When fused to protein 961 [NH₂-ΔG287-961-COOH – sequence shown above], the resulting protein is insoluble and must be denatured and renatured for purification. Following

10 renaturation, around 50% of the protein was found to remain insoluble. The soluble and insoluble proteins were compared, and much better bactericidal titres were obtained with the soluble protein (FCA as adjuvant):

	2996	BZ232	MC58	NGH38	F6124	BZ133
Soluble	65536	128	4096	>2048	>2048	4096
Insoluble	8192	<4	<4	16	n.d.	n.d.

Titres with the insoluble form were, however, improved by using alum adjuvant instead:

Insoluble	32768	128	4096	>2048	>2048	2048
------------------	-------	-----	------	-------	-------	------

15

Example 17 – N-terminal fusions ('hybrids') to 287

Expression of protein 287 as full-length with a C-terminal His-tag, or without its leader peptide but with a C-terminal His-tag, gives fairly low expression levels. Better expression is achieved using a N-terminal GST-fusion.

As an alternative to using GST as an N-terminal fusion partner, 287 was placed at the C-terminus of protein 919 ('919-287'), of protein 953 ('953-287'), and of proteins ORF46.1 ('ORF46.1-287'). In both cases, the leader peptides were deleted, and the hybrids were direct in-frame fusions.

5 To generate the 953-287 hybrid, the leader peptides of the two proteins were omitted by designing the forward primer downstream from the leader of each sequence; the stop codon sequence was omitted in the 953 reverse primer but included in the 287 reverse primer. For the 953 gene, the 5' and the 3' primers used for amplification included a *Nde*I and a *Bam*HI restriction sites respectively, whereas for the amplification of the 287 gene the 5' and the 3' 10 primers included a *Bam*HI and a *Xho*I restriction sites respectively. In this way a sequential directional cloning of the two genes in pET21b+, using *Nde*I-*Bam*HI (to clone the first gene) and subsequently *Bam*HI-*Xho*I (to clone the second gene) could be achieved.

15 The 919-287 hybrid was obtained by cloning the sequence coding for the mature portion of 287 into the *Xho*I site at the 3'-end of the 919-His clone in pET21b+. The primers used for amplification of the 287 gene were designed for introducing a *Sal*II restriction site at the 5'- and a *Xho*I site at the 3'- of the PCR fragment. Since the cohesive ends produced by the *Sal*II and *Xho*I restriction enzymes are compatible, the 287 PCR product digested with *Sal*II-*Xho*I could be inserted in the pET21b-919 clone cleaved with *Xho*I.

20 The ORF46.1-287 hybrid was obtained similarly.

The bactericidal efficacy (homologous strain) of antibodies raised against the hybrid proteins was compared with antibodies raised against simple mixtures of the component antigens:

	Mixture with 287	Hybrid with 287
919	32000	16000
953	8192	8192
ORF46.1	128	8192

Data for bactericidal activity against heterologous MenB strains and against serotypes A and C were also obtained for 919-287 and 953-287:

Strain	919		953		ORF46.1	
	Mixture	Hybrid	Mixture	Hybrid	Mixture	Hybrid
MC58	512	1024	512	1024	-	1024
NGH38	1024	2048	2048	4096	-	4096
BZ232	512	128	1024	16	-	-
MenA (F6124)	512	2048	2048	32	-	1024
MenC (C11)	>2048	n.d.	>2048	n.d.	-	n.d.
MenC (BZ133)	>4096	>8192	>4096	<16	-	2048

Hybrids of ORF46.1 and 919 were also constructed. Best results (four-fold higher titre) were achieved with 919 at the N-terminus.

Hybrids 919-519His, ORF97-225His and 225-ORF97His were also tested. These gave moderate ELISA fitres and bactericidal antibody responses.

5 **Example 18 – the leader peptide from ORF4**

As shown above, the leader peptide of ORF4 can be fused to the mature sequence of other proteins (e.g. proteins 287 and 919). It is able to direct lipidation in *E.coli*.

Example 19 – domains in 564

10 The protein '564' is very large (2073aa), and it is difficult to clone and express it in complete form. To facilitate expression, the protein has been divided into four domains, as shown in figure 8 (according to the MC58 sequence):

Domain	A	B	C	D
Amino Acids	79-360	361-731	732-2044	2045-2073

These domains show the following homologies:

- Domain A shows homology to other bacterial toxins:

15 **gb|AAG03431.1|AE004443_9** probable hemagglutinin [*Pseudomonas aeruginosa*] (38%)
gb|AAC31981.1| (139897) HecA [*Pectobacterium chrysanthemi*] (45%)
emb|CAA36409.1| (X52156) filamentous hemagglutinin [*Bordetella pertussis*] (31%)
gb|AAC79757.1| (AF057695) large supernatant protein1 [*Haemophilus ducreyi*] (26%)
gb|AAA25657.1| (M30186) HpmA precursor [*Proteus mirabilis*] (29%)

20 • Domain B shows no homology, and is specific to 564.

- Domain C shows homology to:

25 **gb|AAF84995.1|AE004032** HA-like secreted protein [*Xylella fastidiosa*] (33%)
gb|AAG05850.1|AE004673 hypothetical protein [*Pseudomonas aeruginosa*] (27%)
gb|AAF68414.1|AF237928 putative FHA [*Pasteurella multocida*] (23%)
gb|AAC79757.1| (AF057695) large supernatant protein1 [*Haemophilus ducreyi*] (23%)
pir|S21010 FHA B precursor [*Bordetella pertussis*] (20%)

- Domain D shows homology to other bacterial toxins:

gb|AAF84995.1|AE004032_14 HA-like secreted protein [Xylella fastidiosa] (29%)

Using the MC58 strain sequence, good intracellular expression of 564ab was obtained in the 5 form of GST-fusions (no purification) and his-tagged protein; this domain-pair was also expressed as a lipoprotein, which showed moderate expression in the outer membrane/supernatant fraction.

The b domain showed moderate intracellular expression when expressed as a his-tagged product (no purification), and good expression as a GST-fusion.

10 The c domain showed good intracellular expression as a GST-fusion, but was insoluble. The d domain showed moderate intracellular expression as a his-tagged product (no purification). The cd protein domain-pair showed moderate intracellular expression (no purification) as a GST-fusion.

Good bactericidal assay titres were observed using the c domain and the bc pair.

15 **Example 20 – the 919 leader peptide**

The 20mer leader peptide from 919 is discussed in example 1 above:

MKKYLFRAAL YGIAAAILAA

As shown in example 1, deletion of this leader improves heterologous expression, as does 20 substitution with the ORF4 leader peptide. The influence of the 919 leader on expression was investigated by fusing the coding sequence to the *PhoC* reporter gene from *Morganella morganii* [Thaller *et al.* (1994) *Microbiology* 140:1341-1350]. The construct was cloned in the pET21-b plasmid between the *NdeI* and *XhoI* sites (Figure 9):

25	1 MKKYLFRAAL YGIAAAILAA AIPAGNDATT KPDLYYLKNE QAIDSLKLLP
	51 PPPEVGSIQF LNDQAMYEKG RMLRNTERGK QAQADADLAA GGVATAFSGA
	101 FGYPITEKDS PELYKLLTNM IEDAGDLATR SAKEHYMIR PFAFYGTETC
	151 NTKDQKKLST NGSYPSGHTS IGWATALVLA EVNPANQDAI LERGYQLGQS
	201 RVICGYHWQS DVDAARIVGS AAVATLHSDP AFQAQLAKAK QEFAQKSQK*

30 The level of expression of *PhoC* from this plasmid is >200-fold lower than that found for the same construct but containing the native *PhoC* signal peptide. The same result was obtained even after substitution of the T7 promoter with the *E.coli* *Plac* promoter. This means that the influence of the 919 leader sequence on expression does not depend on the promoter used.

In order to investigate if the results observed were due to some peculiarity of the 919 signal 35 peptide nucleotide sequence (secondary structure formation, sensitivity to RNAases, *etc.*) or

to protein instability induced by the presence of this signal peptide, a number of mutants were generated. The approach used was a substitution of nucleotides of the 919 signal peptide sequence by cloning synthetic linkers containing degenerate codons. In this way, mutants were obtained with nucleotide and/or amino acid substitutions.

5 Two different linkers were used, designed to produce mutations in two different regions of the 919 signal peptide sequence, in the first 19 base pairs (L1) and between bases 20-36 (S1).

L1: 5' T ATG AAa/g TAc/t c/tTN TTt/c a/cGC GCC GCC CTG TAC GGC ATC GCC GCC
GCC ATC CTC GCC GCC GCG ATC CC 3'
S1: 5' T ATG AAA AAA TAC CTA TTC CGa/g GCN GCN c/tTa/g TAc/t GGc/g ATC GCC
GCC GCC ATC CTC GCC GCC ATC CC 3'

10 The alignment of some of the mutants obtained is given below.

L1 mutants:

15 9L1-a ATGAAAGAAGTACCTTTTCAGCGCCGCC~~~~~
9L1-e ATGAAAAAAATACTTTTCCGCGCCGCC~~~~~
9L1-d ATGAAAAAAATACTTTTCCGCGCCGCC~~~~~
9L1-f ATGAAAAAAATATCTCTTTAGCGCCGCCCTGTACGGCATCGCCGCCATCCTCGCCGCC
919sp ATGAAAAAAATACCTATTCCGCGCCGCCCTGTACGGCATCGCCGCCATCCTCGCCGCC

20 9L1a MKKYLFSA~~~
9L1e MKKYFFRAA~~~
9L1d MKKYFFRAA~~~
9L1f MKKYLFSAALYGIAAAILAA
919sp MKKYLFRAALYGIAAAILAA (i.e. native signal peptide)

S1 mutants:

25 9S1-e ATGAAAAAAATACCTATTTC.....ATCGCCGCCGCCATCCTCGCCGCC
9S1-c ATGAAAAAAATACCTATTCCGAGCTGCCAATACGGCATCGCCGCCATCCTCGCCGCC
9S1-b ATGAAAAAAATACCTATTCCGGGCCGCCAATACGGCATCGCCGCCATCCTCGCCGCC
9S1-i ATGAAAAAAATACCTATTCCGGGCCCTGTACGGATCGCCGCCATCCTCGCCGCC
919sp ATGAAAAAAATACCTATTCCGCGCCGCCCTGTACGGCATCGCCGCCATCCTCGCCGCC

30 9S1e MKKYLF.....IAAAILAA
9S1c MKKYLFRAAQYGIAAAILAA
9S1b MKKYLFRAAQYGIAAAILAA
9S1i MKKYLFRAALYGIAAAILAA
919sp MKKYLFRAALYGIAAAILAA

40 As shown in the sequences alignments, most of the mutants analysed contain in-frame deletions which were unexpectedly produced by the host cells.

Selection of the mutants was performed by transforming *E. coli* BL21(DE3) cells with DNA prepared from a mixture of L1 and S1 mutated clones. Single transformants were screened for high PhoC activity by streaking them onto LB plates containing 100 µg/ml ampicillin, 45 50µg/ml methyl green, 1 mg/ml PDP (phenolphthaleindiphosphate). On this medium PhoC-producing cells become green (Figure 10).

A quantitative analysis of PhoC produced by these mutants was carried out in liquid medium using pNPP as a substrate for PhoC activity. The specific activities measured in cell extracts and supernatants of mutants grown in liquid medium for 0, 30, 90, 180 min. were:

CELL EXTRACTS

	0	30	90	180
control	0,00	0,00	0,00	0,00
9phoC	1,11	1,11	3,33	4,44
9S1e	102,12	111,00	149,85	172,05
9L1a	206,46	111,00	94,35	83,25
9L1d	5,11	4,77	4,00	3,11
9L1f	27,75	94,35	82,14	36,63
9S1b	156,51	111,00	72,15	28,86
9S1c	72,15	33,30	21,09	14,43
9S1i	156,51	83,25	55,50	26,64
phoCwt	194,25	180,93	149,85	142,08

5

SUPERNATANTS

	0	30	90	180
control	0,00	0,00	0,00	0,00
9phoC	0,33	0,00	0,00	0,00
9S1e	0,11	0,22	0,44	0,89
9L1a	4,88	5,99	5,99	7,22
9L1d	0,11	0,11	0,11	0,11
9L1f	0,11	0,22	0,11	0,11
9S1b	1,44	1,44	1,44	1,67
9S1c	0,44	0,78	0,56	0,67
9S1i	0,22	0,44	0,22	0,78
phoCwt	34,41	43,29	87,69	177,60

Some of the mutants produce high amounts of PhoC and in particular, mutant 9L1a can secrete PhoC in the culture medium. This is noteworthy since the signal peptide sequence of 10 this mutant is only 9 amino acids long. This is the shortest signal peptide described to date.

Example 21 – C-terminal deletions of Maf-related proteins

MafB-related proteins include 730, ORF46 and ORF29.

The 730 protein from MC58 has the following sequence:

15 1 VKPLRRLTML LAACAVAAAAA LIQPALAADL AQDPFITDNA QRQHYEPGGK
 51 YHLFGDPRGS VSDRTGKINV IQDYTHQMGN LLIQQANING TIGYHTRFSG
 101 HGHEEHAPFD NHAADSASEE KGNVDEGFTV YRLNWECHEH HPADAYDGPK
 151 GGNYPKPTGA RDEYTYHVNG TARSIKLNPT DTRSIQRQIS DNYSNLGSNF
 201 SDRADEANRK MFEHNAKLDL WGNSMEFING VAAGALNPFFI SAGEALGIGD
 251 ILYGTRYAID KAAMRNIAPL PAEGKFAVIG GLGSVAGFEK NTREAVDRWI
 301 QENPNAAETV EAVFNVAATAA KVAKLAKAAK PGKAAVSGDF ADSYKKKLAL

20

351 SDSARQLYQN AKYREALDIH YEDLIRRKTD GSSKFINGRE IDAVTNDALI
401 QAKRTISAID KPKNFLNQKN RKQIKATIEA ANQQGKRAEF WFKYGVHSQV
451 KSYIESKGGL VKTGLGD*

5 The leader peptide is underlined.

730 shows similar features to ORF46 (see example 8 above):

- as for Orf46, the conservation of the 730 sequence among MenB, MenA and gonococcus is high (>80%) only for the N-terminal portion. The C-terminus, from ~340, is highly divergent.
- 10 – its predicted secondary structure contains a hydrophobic segment spanning the central region of the molecule (aa. 227-247).
- expression of the full-length gene in *E. coli* gives very low yields of protein. Expression from tagged or untagged constructs where the signal peptide sequence has been omitted has a toxic effect on the host cells. In other words, the presence of the full-length mature protein in the cytoplasm is highly toxic for the host cell while its translocation to the periplasm (mediated by the signal peptide) has no detectable effect on cell viability. This “intracellular toxicity” of 730 is particularly high since clones for expression of the leaderless 730 can only be obtained at very low frequency using a *recA* genetic background (*E. coli* strains: HB101 for cloning; HMS174(DE3) for expression).
- 15
- 20 To overcome this toxicity, a similar approach was used for 730 as described in example 8 for ORF46. Four C-terminal truncated forms were obtained, each of which is well expressed. All were obtained from intracellular expression of His-tagged leaderless 730.

Form A consists of the N-terminal hydrophilic region of the mature protein (aa. 28-226). This was purified as a soluble His-tagged product, having a higher-than-expected MW.

- 25 Form B extends to the end of the region conserved between serogroups (aa. 28-340). This was purified as an insoluble His-tagged product.

The C-terminal truncated forms named C1 and C2 were obtained after screening for clones expressing high levels of 730-His clones in strain HMS174(DE3). Briefly, the pET21b plasmid containing the His-tagged sequence coding for the full-length mature 730 protein 30 was used to transform the *recA* strain HMS174(DE3). Transformants were obtained at low frequency which showed two phenotypes: large colonies and very small colonies. Several large and small colonies were analysed for expression of the 730-His clone. Only cells from large colonies over-expressed a protein recognised by anti-730A antibodies. However the

protein over-expressed in different clones showed differences in molecular mass. Sequencing of two of the clones revealed that in both cases integration of an *E. coli* IS sequence had occurred within the sequence coding for the C terminal region of 730. The two integration events have produced in-frame fusion with 1 additional codon in the case of C1, 5 and 12 additional codons in the case of C2 (Figure 11). The resulting "mutant" forms of 730 have the following sequences:

730-C1 (due to an IS1 insertion - figure 11A)

1	MADLAQDPFI TDNAQRQHYE PGGKYHLFGD PRGSVSDRTG KINVIQDYTH
51	<u>QMGNLLIQQQA</u> NINGTIGYHT RFSGHGHEEH APFDNHAADS ASEEKGNVDE
101	GFTVYRLNWE GHEHHPADAY DGPKGGNYPK PTGARDEYTY HVNGTARSIK
151	<u>LNPTDTRSIR</u> QRISDNYSNL GSNFSDRADE ANRKMFEHNA KLDRWGNSME
201	FINGVAAGAL NPFISAGEAL GIGDILYGTR YAIDKAAMRN IAPLPAEGKF
251	AVIGGLGSVA GFEKNTREAV DRWIQENPNA AETVEAVFNV AAAAKVAKLA
301	KAAKPGKAAC SGDFADSYKK KLALSDSARQ LYQNAKYREA LDIHYEDLIR
351	RKTDGSSKFI NGREIDAVTN DALIQAR*

The additional amino acid produced by the insertion is underlined.

730-C2 (due to an IS5 insertion - Figure 11B)

1	MADLAQDPFI TDNAQRQHYE PGGKYHLFGD PRGSVSDRTG KINVIQDYTH
51	<u>QMGNLLIQQQA</u> NINGTIGYHT RFSGHGHEEH APFDNHAADS ASEEKGNVDE
101	GFTVYRLNWE GHEHHPADAY DGPKGGNYPK PTGARDEYTY HVNGTARSIK
151	<u>LNPTDTRSIR</u> QRISDNYSNL GSNFSDRADE ANRKMFEHNA KLDRWGNSME
201	FINGVAAGAL NPFISAGEAL GIGDILYGTR YAIDKAAMRN IAPLPAEGKF
251	AVIGGLGSVA GFEKNTREAV DRWIQENPNA AETVEAVFNV AAAAKVAKLA
301	KAAKPGKAAC SGDFADSYKK KLALSDSARQ LYQNAKYREA <u>LGKVRISEI</u>
351	<u>LLG*</u>

The additional amino acids produced by the insertion are underlined.

In conclusion, intracellular expression of the 730-C1 form gives very high level of protein 30 and has no toxic effect on the host cells, whereas the presence of the native C-terminus is toxic. These data suggest that the "intracellular toxicity" of 730 is associated with the C-terminal 65 amino acids of the protein.

Equivalent truncation of ORF29 to the first 231 or 368 amino acids has been performed, using expression with or without the leader peptide (amino acids 1-26; deletion gives 35 cytoplasmic expression) and with or without a His-tag.

Example 22 – domains in 961

As described in example 9 above, the GST-fusion of 961 was the best-expressed in *E. coli*. To improve expression, the protein was divided into domains (figure 12).

The domains of 961 were designed on the basis of YadA (an adhesin produced by *Yersinia* 40 which has been demonstrated to be an adhesin localized on the bacterial surface that forms

oligomers that generate surface projection [Hoiczyk *et al.* (2000) *EMBO J* 19:5989-99]) and are: leader peptide, head domain, coiled-coil region (stalk), and membrane anchor domain.

These domains were expressed with or without the leader peptide, and optionally fused either to C-terminal His-tag or to N-terminal GST. *E.coli* clones expressing different domains of 961 were analyzed by SDS-PAGE and western blot for the production and localization of the expressed protein, from over-night (o/n) culture or after 3 hours induction with IPTG. The results were:

	Total lysate (Western Blot)	Periplasm (Western Blot)	Supernatant (Western Blot)	OMV SDS-PAGE
961 (o/n)	-	-	-	
961 (IPTG)	+/-	-	-	
961-L (o/n)	+	-	-	+
961-L (IPTG)	+	-	-	+
961c-L (o/n)	-	-	-	
961c-L (IPTG)	+	+	+	
961Δ ₁ -L (o/n)	-	-	-	
961Δ ₁ -L (IPTG)	+	-	-	+

The results show that in *E.coli*:

- 961-L is highly expressed and localized on the outer membrane. By western blot analysis two specific bands have been detected: one at ~45kDa (the predicted molecular weight) and one at ~180kDa, indicating that 961-L can form oligomers. Additionally, these aggregates are more expressed in the over-night culture (without IPTG induction). OMV preparations of this clone were used to immunize mice and serum was obtained. Using overnight culture (predominantly by oligomeric form) the serum was bactericidal; the IPTG-induced culture (predominantly monomeric) was not bactericidal.
- 961Δ₁-L (with a partial deletion in the anchor region) is highly expressed and localized on the outer membrane, but does not form oligomers;
- the 961c-L (without the anchor region) is produced in soluble form and exported in the supernatant.

Titres in ELISA and in the serum bactericidal assay using His-fusions were as follows:

	ELISA	Bactericidal
961a (aa 24-268)	24397	4096

961b (aa 269-405)	7763	64
961c-L	29770	8192
961c (2996)	30774	>65536
961c (MC58)	33437	16384
961d	26069	>65536

E.coli clones expressing different forms of 961 (961, 961-L, 961 Δ 1-L and 961c-L) were used to investigate if the 961 is an adhesin (c.f. YadA). An adhesion assay was performed using (a) the human epithelial cells and (b) *E.coli* clones after either over-night culture or three hours IPTG induction. 961-L grown over-night (961 Δ 1-L) and IPTG-induced 961c-L (the clones expressing protein on surface) adhere to human epithelial cells.

961c was also used in hybrid proteins (see above). As 961 and its domain variants direct efficient expression, they are ideally suited as the N-terminal portion of a hybrid protein.

Example 23 – further hybrids

Further hybrid proteins of the invention are shown below (see also Figure 14). These are advantageous when compared to the individual proteins:

ORF46.1-741		
1	ATGTCAGATT	TGGCAAACGA
51	GCATTTGAA	CCCGACGGGA
101	TTGCCGAGCG	CAGCGGCCAT
151	TTGGGCAACC	TGATGATTCA
201	CATTGTCGCG	ACAGGGCGGC
251	ACCATCCGAT	ATTAAAGGAA
301	TTTAGCCTT	ATATCGGCTA
351	CGGCTATGAC	ACGGGCACGA
401	GGGATATATA	AGTGTAGAAG
451	ACACCTGACCG	CCGGTACGGA
501	AACTGCGGT	TCTGATGAAG
551	CAATGCCGAT	GGGGACCGA
601	CCACCCGATA	TACGAACACC
651	TTCAACGGCA	ATCCCCTCGA
701	CTGCAGATAT	AAAGGGCGGA
751	GGGCTATGAC	TCCCTCGACT
801	GGGATATATA	AAAGGGCGCA
851	ACCATCCGCG	TCCCTCGACT
901	GGGCTATGAC	AAAGGGCGCA
951	GGGCTATGAC	TCCCTCGACT
1001	GGGCTATGAC	AAAGGGCGCA
1051	GGGCTATGAC	TCCCTCGACT
1101	GGGCTATGAC	AAAGGGCGCA
1151	GGGCTATGAC	TCCCTCGACT
1201	GGGCTATGAC	AAAGGGCGCA
1251	GGGCTATGAC	TCCCTCGACT
1301	GGGCTATGAC	AAAGGGCGCA
1351	GGGCTATGAC	TCCCTCGACT
1401	GGGCTATGAC	AAAGGGCGCA
1451	GGGCTATGAC	TCCCTCGACT
1501	GGGCTATGAC	AAAGGGCGCA
1551	GGGCTATGAC	TCCCTCGACT
1601	GGGCTATGAC	AAAGGGCGCA

1651	GACAAGCTTC	CCGAAGGCGG	CAGGGCGACA	TATCGCGGGA	CGGCCTTCGG	
1701	TTCAGACGAT	GCCGGCGGAA	AACTGACCTA	CACCATAGAT	TCGCGGCCA	
1751	AGCAGGGAAA	CGGAAATAC	GAACATTGAA	ATCGCCAGA	ACTCAATGTC	
5	1801	GACTTGGCCG	CCGGCGATAT	CAAGCCGGAT	GGAAAACGCC	ATGCCGTCAT
1851	CAGCGGTTCC	GTCCTTTACA	ACCAAGCCGA	GAAAGGCAGT	TACTCCCTCG	
1901	GTATCTTGG	CGGAAAAGCC	CAGGAAGTTG	CCGGCAGCGC	GGAAGTGAAA	
1951	ACCGTAAACG	GCATACGCCA	TATCGCCCTT	GGCGCCAAGC	AACTCGAGCA	
2001	CCACCAACAC	CACCACTGA				
10	1	MSDLANDSF	RQVLDRQHFE	PDGKYHLFGS	RGELAERSGH	IGLGKIQSHQ
	51	LGNLMIQQAA	IKGNIGYIVR	FSDHGHEVHS	PFDNHASHSD	SDEAGSPVDG
	101	FSLYRIHWG	YEHHPADGYD	GPQGGGYPAP	KGARDIYSYD	IKGVAQNIRL
	151	NLTDNRSTGQ	RLADRFHNAG	SMLTQGVGPDG	FKRATRYSPE	LDRSGNAAEA
15	201	FNGTADIVKN	IIIGAAGEIVG	AGDAVQGISE	GSNTIAVMHGL	GLLSTENKMA
	251	RINDLADMAQ	LKDYAAAIR	DWAVQNPNAA	QGIEAVSNIF	MAAIPIKGIG
	301	AVRGKYGLGG	ITAHPIKRSQ	MGAIALPKGK	SAVDNFADA	AYAKYPSPYH
	351	SRNIRSNLHQ	RYGKENITSS	TVPPSNGKNV	KLADQRHPKT	GVPFDGKGFP
	401	NFEKHVKYDT	GSGGGGVAAD	IGAGLADALT	APLDHKDKGL	QLSLTLDQSVR
20	451	KNEKLKLAAQ	GAEKTYGNQD	SLNTGKLKND	KVSRFDFIRO	IEVDGQLIITL
	501	ESGEFQVYKQ	SHSALTAFQT	EQIQDSEHSG	KMVAKRQFRI	GDIAGEHTSF
	551	DKLPEGGRAT	YRGTAFGSDD	AGGKLTYTID	FAAKQGNGKI	EHLKSPELNV
	601	DLAAADIKPD	GKRHAVISGS	VLYNQAEKGS	YSLGIFGGKA	QEVAWSAEVK
	651	TVNGIRHIGL	AAKQLEHHHH	HH*		
25						
	ORF46.1-961					
	1	ATGTCAGATT	TGGCAAACGA	TTCTTTTATC	CGGCAGGTT	TCGACCGTCA
	51	GCATTTCGAA	CCCGACGGG	AATACCACCT	ATTCGGCAGC	AGGGGGGAAC
30	101	TTGCCGAGCG	CAGCGGCCAT	ATCGGATTGG	AAAAAATACA	AAGCCATCAG
	151	TTGGGCAACC	TGATGATTCA	ACAGGCAGGC	ATTAAAGGAA	ATATCGGCTA
	201	CATTGTCGC	TTTTCGATC	ACGGGCACGA	AGTCCATTCC	CCCTTCGACA
	251	ACCATGCCTC	ACATTCCGAT	TCTGATGAAG	CCGGTAGTC	CGTTGACGGA
	301	TTTACGCTTT	ACCGCATCCA	TTGGGACGGA	TACGAACACC	ATCCCAGCGA
35	351	CGGCTATGAC	GGGCCACAGG	GGGGCGGCTA	TCCCCTC	AAAGGCGCGA
	401	GGGATATATA	CAGCTACCGAC	ATAAAAGGCG	TTGCCCCAAA	TATCCCGCTC
	451	AACCTGACCG	ACAACCGCAG	CACCGGACAA	CGGCTTGCCG	ACCGTTTCCA
	501	CAATGCCGGT	AGTATGCTGA	CGCAAGGAGT	AGGCGACGGA	TCACAAACCGC
	551	CCACCCGATA	CAGCCCGAG	CTGGACAGAT	CGGGCAATGC	CGCCGAAGCC
40	601	TTCACAGGCA	CTGCAGATAT	CGTTAAAAC	ATCATCGGCC	CGGCAGGAGA
	651	AATTGTCGGC	GCAGGCATG	CCGTGCAGGG	CATAAGCGAA	GGCTAAACA
	701	TTGCTGTCAT	GCACGGCTT	GGTCTGCTT	CCACCGAAAA	CAAGATGGCG
	751	CGCATCAACG	ATTTGGCAGA	TATGGCGCAA	CTCAAAGACT	ATGCCGCAGC
	801	AGCCATCCGC	GATTGGGAG	TCCAAAACCC	CAATGCCGCA	CAAGGCATAG
45	851	AAGCCGTCAG	CAATATCTT	ATGGCAGCCA	TCCCCATCAA	AGGGATTTGGA
	901	GCTGTCGGG	GGAAAATACGG	CTTGGGCGGC	ATCACGGCAC	ATCCTATCAA
	951	CGGGTCGCGAG	ATGGGCGCGA	TCGCATTGCC	GGAAAGGAAA	TCCGCGTCA
	1001	GGCACAATT	TGCCGATGCG	GCATACGCCA	AATACCGTC	CCCTTACCAT
	1051	TCCCGAAATA	TCCGTTCAA	CTTGGAGCAG	CGTTACGGCA	AAGAAAACAT
50	1101	CACCTCCTCA	ACCGTCCCG	CGTCAAACGG	AAAAAATGTC	AAACTGGCAG
	1151	ACCAACGCCA	CCCGAAGACA	GGCGTACCGT	TTGACGGTAA	AGGGTTTCCG
	1201	AATTTCGAGA	AGCACGTGAA	ATATGATACG	GGATCCGGAG	GAGGAGGAGC
	1251	CACAAACGAC	GACGATGTTA	AAAAAGCTGC	CACTGTGGC	ATTGCTGCTG
	1301	CCTACAACAA	TGGCAAGAA	ATCAACGGTT	TCAAAGCTGG	AGAGACCATC
55	1351	TACGACATTG	ATGAAGACGG	CACAAATTAC	AAAAAAGAGC	CAACTGCAGC
	1401	CGATGTTGAA	GCCGACGACT	TTAAAGGTCT	GGGCTGAAA	AAAGTCGTGA
	1451	CTAACCTGAC	CAAAACCGTC	AATGAAAACA	AACAAACGT	CGATGCCAA
	1501	GTAAAAGCTG	CAGAACTGTA	AATAGAAAAG	TTAACACCA	AGTTAGCAGA
	1551	CACTGATGCC	GCTTTAGCAG	ATACTGATGC	CGCTCTGGAT	GCAACCCACCA
60	1601	ACGCCCTGAA	TAAATTGGGA	GAAAATATAA	CGACATTGTC	TGAAGAGACT
	1651	AAGACAAATA	TCGTAAAAAT	TGATGAAAAA	TTAGAAGCCG	TGGCTGATAC
	1701	CGTCGACAAG	CATGCCGAAG	CATTCAACGA	TATCGCCGAT	TCATTGGATG
	1751	AAACCAACAC	TAAGGCAGAC	GAAGCCGTCA	AAACCGCCAA	TGAAGCCAAA
	1801	CAGACGGCCG	AAGAAACCAA	ACAAAACGTC	GATGCCAAAG	TAAAAGCTGC
	1851	AGAAAATGCA	GCAGGCAAAAG	CCGAAGCTGC	CGCTGGCACA	GCTAATACTG
65	1901	CAGCCGACAA	GGCCGAAGCT	GTCGCTGCAA	AGTTACCGA	CATCAAAGCT
	1951	GATATCGCTA	CGAACAAAGA	TAATATTGCT	AAAAAAGCAA	ACAGTGCCGA
	2001	CGTGTACACC	AGAGAAGAGT	CTGACAGCAA	ATTGATGGTC	

2051	TGAACGCTAC	TACCGAAAAA	TTGGACACAC	GCTTGGCTTC	TGCTGAAAAA	
2101	TCCATTGCCG	ATCACGATAC	TCGCCTGAAC	GGTTGGATA	AAACAGTGTGTC	
2151	AGACCTGCGC	AAAGAAACCC	GCCAAGGCCT	TGCAGAACAA	GCCGCGCTCT	
2201	CCGGTCTGTT	CCAACCTAC	AACGTGGGTC	GTTCAATGT	AACGGCTGCA	
5	2251	GTCGGCGGCT	ACAAATCCGA	ATCGGCAGTC	GCCATCGGTAA	CCGGCTTCCG
2301	CTTTACCGAA	AACTTTGGCG	CCAAAGCAGG	CGTGGCAGTC	GGCACTTCGT	
2351	CCGGTTCTTC	CGCAGCTAC	CATGTCGGCG	TCAATTACGA	GTGGCTCGAG	
2401	CACCACCACC	ACCACCACTG	A			
10	1	MSDLANDSFI	RQVLDQRHFE	PDGKYHLFGS	RGELAERSGH	IGLGKIQSHQ
	51	LGMLMTQQAA	IKGNIGYIVR	FSDHGHEVHS	PFDNHASHSD	SDEAGSPVDG
	101	FSLYRIHWDG	YEHHPADGYD	GPQGGGYPAP	KGARDIYSYD	IKGVAQNIRL
	151	NLTDRNRTGQ	RLADRFHNAG	SMLTQGVGDG	FKRATRYSPE	LDRSGNAAEA
15	201	FNGTADIVKN	IIIGAAGEIVG	AGDAVQGISE	GSNIAVMHGL	GLLSTENKMA
	251	RINDLADMAQ	LKDYAAAIR	DWAVQNPNAA	QGIEAVSNIF	MAAIPIKGIG
	301	AVRGKYGLGG	ITAHPIKRSQ	MGAIALPKGK	SAVSDNFADA	AYAKYPSPYH
	351	SRNIRSNLEQ	RYGKENITSS	TVPPSNGKNV	KLADQRHPKT	GVPFDGKGFP
	401	NFEKHVKYDT	GSGGGGATND	DDVKKAAATVA	IAAAYNNGQE	INGFKAGETI
20	451	YDIDEDGTIT	KKDATAADVE	ADDFKGLGLK	KVVTNLTKTV	NENKQNDAK
	501	VKAAESEIEK	LTTKLAJDTDA	ALADTDAALD	ATTNALNKLG	ENITTFAEET
	551	KTNIVKIDEK	LEAVADTVDK	HAEAFNDIAD	SLDETNTKAD	BAVKTANEAK
	601	QTAEETKQNV	DAVKAAETA	AGKAEAAAAGT	ANTAADKAEA	VAAKVTDIKA
	651	DIATNKNJIA	KKANSADVYT	REESDSKFVR	IDGLNATTEK	LDTRLASAEK
25	701	SIADHDTRLN	GLDKTVSDLR	KETRQGLAEQ	AALSGLFQPY	NVGRFNVTAA
	751	VGGYKSESACV	AITGTGFRFTE	NFAAKAGVAV	GTSSGSSAA	HVGVNYEWLE
	801	HHHHHH*				

ORF46.1-961c

30	1	ATGTCAGATT	TGGCAAACGA	TTCTTTTATC	CGGCAGGTTTC	TCGACCGTCA
	51	GCAATTTCGAA	CCCGACGGGA	AATACCACCT	ATTCCGGCAGC	AGGGGGGAAC
	101	TTGGCGAGCG	CAGCGGCCAT	ATCGGATTGG	AAAAAAATACA	AAGCCATCAG
	151	TTGGGCAACC	TGATGATTCA	ACAGGGCGGCC	ATTAAAGGAA	ATATCGGCTA
	201	CATTGTCCGC	TTTTCCGATC	ACGGGCACGA	AGTCCATTCC	CCCTTCGACA
35	251	ACCATGCCTC	ACATTCCGAT	TCTGATGAAG	CCGGTAGTGTCC	CGTTGACCGA
	301	TTTAGCCTTT	ACCGCATCCA	TTGGGACGGGA	TACGAACACC	ATCCCGCCGA
	351	CGGCTATGAC	GGGCCACAGG	CGGGCGGCTA	TCCCGCTCCC	AAAGGCGCGA
	401	GGGATATATA	CAGCTACCGAC	ATAAAAGGCG	TTGCCAAAAA	TATCCGCCTC
40	451	AACCTGACCG	ACAACCGCAG	CACCGGACAA	CGGCTTGCCG	ACCGTTTCCA
	501	CAATGCCGGT	AGTATGCTGA	CGCAAGGAGT	AGGCGACCGA	TTCAACACGCG
	551	CCACCCGATA	CAGCCCCCGAG	CTGGACAGAT	CGGGCAATGC	CGCCGAAGCC
	601	TTCAACGGCA	CTGCAGATAT	CGTTAAAAAC	ATCATCGGCG	CGGCAGGAGA
	651	AATTGTCGGC	GCAGGGCGATG	CCGTGCAGGG	CATAAGCGAA	GGCTCAAACA
45	701	TTGCTGTGAT	GCACGGCTTG	GGTCTGCTTT	CCACCGAAAAA	CAAGATGGCG
	751	CGCATCAACG	ATTTGGCAGA	TATGGCGAA	CTCAAAGACT	ATGCCGCAGC
	801	AGCCATCCGC	GATTGGGCAG	TCCAAAACCC	CAATGCCGCA	CAAGGCATAG
	851	AAGCCGTCAG	CAATATCTT	ATGGCAGCCA	TCCCCATCAA	AGGGATTGGA
	901	GCTGTTCGGG	AAAATACGG	CTTGGCGGC	ATCACGGCAC	ATCCTATCAA
50	951	GGGGTCGCAG	ATGGGCGCGA	TCGCATTGCC	GAAGGGAAA	TCCGCGTCA
	1001	GGGACAATT	TGGCGATGCG	GCATACGCCA	AATACCCGTC	CCCTTACCAT
	1051	TCCCGAAATA	TCCGTTCAAA	CTTGGAGCAG	CGTTACGGCA	AAGAAAACAT
	1101	CACCTCCTCA	ACCGTGC CGC	CGTCAAAACGG	AAACTGGCAG	
	1151	ACCAACGCCA	CCCGAAGACA	GGCGTACCGT	TTGACGGTAA	AGGGTTTCCG
55	1201	AATTGTCAGA	AGCACGTGAA	ATATGATACG	GGATCCGGAG	GAGGAGGAGC
	1251	CACAAACGAC	GACGATGTTA	AAAAAGCTGC	CACTGTGGCC	ATTGCTGCTG
	1301	CCTACAACAA	TGGCCAAGAA	ATCAACGGTT	TCAAAGCTGG	AGAGACCATC
	1351	TACGACATTG	ATGAAGACGG	CACAATTACC	AAAAAAAGACG	CAACTGCAGC
	1401	CGATGTTGAA	GCCGACGACT	TTAAAGGTCT	GGGTCTGAAA	AAAGTCGTGA
60	1451	CTAACCTGAC	CAAAACCGTC	AATGAAAACA	AACAAAACGT	CGATGCCAAA
	1501	GTAAGGCTG	CAGAACATG	AATAGAAAAG	TTAACAAACCA	AGTTAGCAGA
	1551	CACTGATGCC	GCTTAGCGAG	ATACTGATGC	CGCTCTGGAT	GCAACCACCA
	1601	ACGCCATTGAA	TAATTGGGA	GAAAATATAA	CGACATTTCG	TGAAGAGACT
	1651	AAGACAAATA	TCGTAAAAAT	TGATGAAAAA	TTAGAAGCCG	TGGCTGATAC
	1701	CGTCGACAAG	CATGCCGAAG	CATTCAACGA	TATGCCGAT	TCATTGGATG
65	1751	AAACCAACAC	TAAGGCAGAC	GAAGCCGTCA	AAACGCCAA	TGAAGCCAAA
	1801	CAGACGGCCG	AAGAAACCAA	ACAAAACGTC	GATGCCAAAG	AAAAGCTGC
	1851	AGAAACTGCA	GCAGGCAAAG	CCGAAGCTGC	CGCTGGCACA	GCTAATACTG

-66-

1901	CAGCCGACAA	GGCCGAAGCT	GTCGCTGCAA	AAGTTACCGA	CATCAAAGCT	
1951	GATATCGCTA	CGAACAAAGA	TAATATTGCT	AAAAAAAGCAA	ACAGTGCAG	
2001	CGTGTACACC	AGAGAAAGAGT	CTGACAGCAA	ATTGTCAGA	ATTGATGGTC	
2051	TGAAACGCTAC	TACCGAAAAA	TTGGACACAC	GCTTGGCTTC	TGCTGAAAAA	
5	2101	TCCATTGCCG	ATCACGATAC	TCGCCTGAAC	GGTTTGGATA	AAACAGTGT
2151	AGACCTGCGC	AAAGAAACCC	GCCAAGGCCT	TGCAGAACAA	GCCGCGCTCT	
2201	CCGGTCTGTT	CCAACCTTAC	AACGTGGGT	TCGAGCACCA	CCACCACAC	
2251	CACTGA					
10	1	MSDLANDSF	RQVLDRQHFE	PDGKYHLFGS	RGELAERSGH	IGLCKIQSHQ
	51	LGNLMIQQAA	IKGNIGYIVR	FSDHGHEVHS	PFDNHASHSD	SDEAGSPVDG
	101	FSLYRIHWDG	YEHHPADGYD	GPQGGGYPAP	KGARDIYSYD	IKGVAQNIRL
	151	NLTDDNRSTGQ	RLADRHNAG	SMLTQGVGDG	FKRATRYSPE	LDRSGNAAEA
15	201	FNGTADIVKN	IIIGAAGEIVG	AGDAVQGISE	GSNIAVMHGL	CLLSTENKMA
	251	RINDLADMAQ	LKDYAAAIR	DWAVQNPNAA	QGIEAVSNIF	MAAIPIKGIG
	301	AVRKGYGLGG	ITAHPIKRSQ	MGAIALPKGK	SAVSDFNADA	AYAKYPSPYH
	351	SRNIRSNLQE	RYGKENITSS	TVPPSNGKNV	KLADQRHFKT	GVPPFDGKGFP
	401	NFEKHVKYD	TGSGGGATND	DDVVKAAATVA	IAAAAYNNQ	INGFKAGETI
20	451	YDIDEDGTIT	KKDATAADVE	ADDFKGLGLK	KVVTNLTKTV	NEBKQNVDAK
	501	VKAAESEIEK	LTTKLADTDA	ALADTDAALD	ATTNALNKLG	ENITTFAEET
	551	KTNIVKIDEK	LEAVADTVDK	HAEAFNDIAD	SLDETNKAD	BAVKTANEAK
	601	QTAEETKQNV	DAKVKAETA	AGKAEAAAGT	ANTAADKAEA	VAAKVTDIKA
	651	DIATNKDNIA	KKANSADVYT	REESDSKFVR	IDGLNATTEK	LDTRILASAEK
25	701	SIADHDTRLN	GLDKTVSDLR	KETRQGLAEQ	AALSGLFQPY	NVGLEHHHHH
	751	H*				

961-ORF46.1

30	1	ATGGCCACAA	ACGACGACGA	TGTTAAAAAA	GCTGCCACTG	TGGCCATTGC
	51	TGCTGCCTAC	AAACATGGCC	AAGAAATCAA	CGGTTTCAA	GCTGGAGAGA
	101	CCATCTACGA	CATTGATGAA	GACGGCACAA	TTACCAAAA	AGACGCACT
	151	GCAGCCGATG	TTGAAGCCGA	CGACTTTAAA	GGTCTGGGT	TGAAAAAAAGT
	201	CGTGAACAA	CTGACCAAAA	CCGTCATGA	AAACAAACAA	AACGTCGATG
35	251	CCAAAGTAAA	AGCTGCAGAA	TCTGAAATAG	AAAAGTTAAC	AACCAAGTTA
	301	GCAGACACTG	ATGCCCTTT	ACGAGATACT	GATCCCGCTC	TGGATGCAAC
	351	CACCAACGCC	TTGAATAAAAT	TGGGAGAAAAA	TATAACGACA	TTTGCTGAAG
	401	AGACTAAGAC	AAATATCGTA	AAAATTGATG	AAAAATTAGA	AGCCGTGGCT
	451	GATACCGTCG	ACAAGCATGC	CGAAGCATTC	AACGATATCG	CCGATTTCATT
40	501	GGATGAAACC	AACACTAAGG	CAGACGAAGC	CGTCAAAACC	GCCAATGAAG
	551	CCAAACAGAC	GGCCGAAGAA	ACCAAACAA	ACGTCGATGC	CAAAGTAAAAA
	601	GCTGCAGAAA	CTGCAGCAGG	CAAAGCCGAA	GCTGCCGCTG	GCACAGCTAA
	651	TACTGCAGCC	GACAAGGCCG	AAGCTGTCGC	TGCAAAAGTT	ACCGACATCA
	701	AAGCTGATAT	CGCTACCAAA	AAAGATAATA	TTGCTAAAAA	AGCAAACAGT
45	751	GCCGACGTGT	ACACCAAGAGA	AGAGTCTGAC	ACGAAATTG	TCAGAATTGA
	801	TGGTCTGAAC	GCTACTACCG	AAAATTGGA	CACACGCTTG	GCTTCTGCTG
	851	AAAAATCCAT	TGCGCATCAC	GATACTCGCC	TGAACGGTTT	GGATAAAAACA
	901	GTGTCAGACC	TGCGCAAAGA	AACCCGCCAA	GGCCTTGCAG	AACAAGCCGC
	951	GCTCTCCGGT	CTGTTCCAAC	CTTACAAACGT	GGGTGGTTTC	AATGTAACGG
50	1001	CTGCAGTCGG	CGGCTACAAA	TCCGAATCGG	CAGTCGCCAT	CGGTACCGGC
	1051	TTCGGCTTTA	CCGAAAACCTT	TGCCGCCAAA	GCAGGGCTGG	CAGTCGGCAC
	1101	TTCGTCGGGT	TCITTCGGCAG	CCTTACCATGT	CGGGCGTCAAT	TACGAGTGGG
	1151	GATCCGGAGG	AGGAGGATCA	GATTTGGCAA	ACGATTCTT	TATCCGGCAG
	1201	GTTCTCGACC	GTCAGCAATT	CGAACCCGAC	GGGAATATACC	ACCTATTGCG
55	1251	CAGCAGGGGG	GAACCTTCCG	AGCCGAGCGG	CCATATCGGA	TTGGGAAAAAA
	1301	TACAAGGCCA	TCAGTGGGC	AACCTGATGA	TTCAACAGGC	GGCCATTAAA
	1351	GAAAATATCG	GCTACATTGT	CCGCTTTTC	GATCACGGGC	ACGAAGTCCA
	1401	TTCCCCCTTC	GACAACCATG	CCTCACATTC	CGATTCTGAT	GAAGCCGTA
	1451	GTCCCGTTGA	CGGATTTCAG	CTTACCGCA	TCCATTGGGA	CGGATACGAA
60	1501	CACCATCCCG	CCGACGCTA	TGACGGGCCA	CAGGGCGGG	GCTATCCCGC
	1551	TCCCAAAGGC	GCGAGGGATA	TATACAGCTA	CGACATAAAA	GGCGTTGCC
	1601	AAAATATCCG	CCTCAACCTG	ACCGACAACC	GCAGCACCGG	ACAACGGCTT
	1651	GCCGACCGTT	TCCACAAATGC	CGGTAGTATG	CTGACGCAAG	GAGTAGGCAG
	1701	CGGATTCAA	CGCGCCACCC	GATACAGCCC	CGAGCTGGAC	AGATCGGGCA
65	1751	ATGCGCCCGA	AGCCTTCAC	GGCACTGCAG	ATATCGTTAA	AAACATCATC
	1801	GGCGCGGCAG	GAGAAATTGT	CGGCGCAGGC	GATGCCGTGC	AGGGCATAAG
	1851	CGAAGGCTCA	AACATTGCTG	TCATGCACGG	CTTGGGTCTG	CTTTCCACCG
	1901	AAAACAAGAT	GGCGCGCATC	AACGATTGG	CAGATATGGC	GCAACTCAAA

1951	GACTATGCCG	CAGCAGCCAT	CCGCGATTGG	GCAGTCCAAA	ACCCCAATGC
2001	CGCACAAAGC	ATAGAAGCCG	TCAGCAATAT	CTTTATGGCA	GCCATCCCCA
2051	TCAAAGGGAT	TGGAGCTGTT	CGGGGAAAAT	ACGGCTTGGG	CGGCATCACG
2101	GCACATCCTA	TCAAGCGTC	GCAGATGGGC	CGGATCGCAT	TGCCGAAAGG
2151	GAAATCCGCC	GTCAGCGACA	ATTGTGCCGA	TGCGGCATAC	GCCAAATAACC
2201	CGTCCCCTTA	CCATTCCCGA	AATATCCGTT	CAAACTTGGA	GCAGCGTTAC
2251	GGCAAAGAAA	ACATCACCTC	CTCAACCGTG	CCGCCGTCAA	ACGGCAAAAAA
2301	TGTCAAACGT	GCAGACCAAC	GCCACCCGAA	GACAGGGCGTA	CCGTTTGACG
2351	GTAAAGGGTT	TCCGAATTT	GAGAAGCACG	TGAAATATGA	TACGCTCGAG
2401	CACCAACCAC	ACCACCACTG	A		
1	MATNDDDVKK	AATVAIAAAAY	NNGQEINGFK	AGETIYDIDE	DGFIKKDAT
51	AADVEADDK	GLGLKKVVTN	LTKTVNENKQ	NVDAKVKAAB	SEIEKLTTKL
101	ADTDAAALADT	DAALDATDNA	LNKLGENITT	FAEETKTNIV	KIDEKLEAVA
151	DTVDKHAEEAF	NDIADSLDET	NTKADEAVKT	ANEAKQTABEE	TKQNVDAKVK
201	AAETAAGKAE	AAAGTANTAA	DKAEAVAAKV	TDIKADIATN	KDNIAKKANS
251	ADVYTREESD	SKFVRIDLGN	ATTEKLDTRL	ASAEKSIADH	DTRLNGLDKT
301	VSDLRKETRQ	GLAEQAALSG	LFQPYNVGRF	NVTAAVGGYK	SESAVAIGTG
351	FRFTENFAAK	AGVAVGTSSG	SSAAVHVGVN	YEWSGGGGS	DILANDSFIRQ
401	VLDQRHFEPD	GKYHLFGSRG	ELAERSGHIG	LGKIQSHQLG	NLMIQQAAIK
451	GNIGYIVRFS	DHGHEVHSPF	DNHASHSDSD	EAGSPVDGFS	LYRIHWGDE
501	HHPADGYDGP	QGGGYPAPKG	ARDIYSYDIK	GVAQNIRLNL	TDNRSTGQRL
551	ADRFLHNAGSM	LTQGVGDGF	RATRYSPELD	RSGNAAEAFN	GTADIVKNII
601	GAAGEIVGAG	DAVQGISEGS	NIAVMHGLGL	LSTENKMARI	NDLADMAQLK
651	DYAAAAAIRDW	AVQNPNAOQ	IEAVSNIFMA	AIPKIGIGAV	RGKYGLGGIT
701	AHPIKRSQLM	AIALPKGKSA	VSDNFADAAY	AKYPSPYHSR	NIRSNLEQRY
751	GKENITSSTV	PPSNGKNVKL	ADQRHPKTGV	PFDGKGFPNF	EKHVKYDTLE
801	HHHHHH*				
30	961-741				
1	ATGGCCACAA	ACGACGACGA	TGTTAAAAAA	GCTGCCACTG	TGGCCATTGC
51	TGCTGCCTAC	AACAATGCC	AAGAAATCAA	CGGTTTCAAA	GCTGGAGAGA
101	CCATCTACGA	CATTGATGAA	GACGGCACAA	TTACACAAA	AGACGCAACT
151	GCAGCCGATG	TTGAAGCCGA	CGACCTTAA	GGTCTGGGT	TGAAAAAAAGT
201	CGTGACTAAC	CTGACCAAA	CCGTCATGA	AAACAAACAA	ACGTCGATG
251	CCAAAGTAAA	AGCTGCAGAA	TCTGAAATAG	AAAAGTTAAC	ACCAAGTTA
301	GCAGACACTG	ATGCCGCTT	AGCAGATACT	GATGCCGCTC	TGGATGCAAC
351	CACCAACGCC	TTGAATAAA	TGGGAGAAAA	TATAACGACA	TTTGCTGAAG
401	AGACTAAGAC	AAATATCGTA	AAAATTGATG	AAAAATTAGA	AGCCGTGGCT
451	GATACCGTCG	ACAAGCATGC	CGAAGCATTG	AACGATATCG	CCGATTCAATT
501	GGATGAAACC	AAACACTAAGG	CAGACGAAGC	CGTCAAAACC	GCCAATGAAG
551	CCAAACAGAC	GGCCGAAGAA	ACCAACACAA	ACGTCGATGC	CAAAGTAAAA
601	GCTGCAGAAA	CTGCAGCAGG	CAAAGCCGAA	GCTGCCGCTG	GCACAGCTAA
651	TACTGCAGCC	GACAAGCCG	AACTGTCGCG	TGCAAAAGTT	ACCGACATCA
701	AAGCTGATAT	CGCTACGAAC	AAAGATAATA	TTGCTAAAAAA	AGCAAAACAGT
751	GCCGACGTGT	ACACCAGAGA	AGAGTCTGAC	AGCAAATTG	TCAGAAATTGA
801	TGGTCTGAAC	GCTACTACCG	AAAATTGGA	CACACGCTT	GCTTCTGCTG
851	AAAAATCCAT	TGCCGATCAC	GATACTGCC	TGAACGGTTT	GGATAAAAACA
901	GTGTCAGACC	TGCGCAAAGA	AACCCGCCAA	GGCCTTGCG	AACAAGCCGC
951	GCTCTCCGGT	CTGTTCCAAAC	CTTACAACGT	GGGTCGGITC	AATGTAACGG
1001	CTGCGAGTCGG	CGGCTACAA	TCCGAAATCGG	CACTGCCCAT	CGGTACCGGC
1051	TTCCGCTTA	CCGAAAACCT	TGCCGCCAA	GCAGGGGTGG	CAGTCGGCAC
1101	TTCCGTCGGT	TCTTCCCGAG	CCTACCATGT	GGGGCGTCAAT	TACGACTGGG
1151	GATCCGGAGG	GGGTGGTGT	GCCGCCAGAC	TCGGTGCAGGG	GCTTGGCGAT
1201	GCACTAACCG	CACCGCTCGA	CCATAAAGAC	AAAGGTTTGC	AGTCTTTGAC
1251	GCTGGATCAG	TCCGTCAGGA	AAAACGAGAA	ACTGAACCTG	GCGGCACAAG
1301	GTGCGGAAAA	AACTTATGGA	AACGGTGACA	GCCTCAATAC	GGGAAATTG
1351	AAGAACGACA	AGGTCAAGCCG	TTTCGACTTT	ATCCGCCAA	TCGAAGTGG
1401	CGGGCAGCTC	ATTACCTGG	AGAGTGGAGA	GTTCCAAGTA	TACAAACAAA
1451	GCCATTCCGC	CTTAACCGCC	TTTCAGACCG	AGCAAATACA	AGATTCCGGAG
1501	CATTCCGGGA	AGATGGTGC	GAAACGCCAG	TTCAGAATCG	GCGACATAGC
1551	GGGCGAACAT	ACATCTT	ACAAGCTTCC	CGAAGGGCGC	AGGGCGACAT
1601	ATCGCGGGAC	GGCGTTCCGT	TCAGACGATG	CCGGCGGAAA	ACTGACCTAC
1651	ACCATAGATT	TCGCGCCAA	GCAGGGAAAC	GGCAAATCG	AAACATTGAA
1701	ATCGCCAGAA	CTCAATGTG	ACCTGGCCCG	CGCCGATATC	AAGCCGGATG
1751	GAAAACGCCA	TGCCGTCATC	AGCGGTTCCG	TCCCTTACAA	CCAAGCCGAG
1801	AAAGGCAGTT	ACTCCCTCGG	TATCTTGGC	GGAAAAGCCC	AGGAAGTTGC

1851 CGGCAGCGCG GAAGTGAAAA CCGTAAACGG CATAGCCAT ATCGGCCTTG
 1901 CCGCCAAGCA ACTCGAGCAC CACCACCA ACCACTGA

 5 1 MATNDDDVKK AATVIAAAAY NNGQEINGFK AGETIYDIDE DGTITKKDAT
 51 AADVEADDFK GLGLKKVVTN LTKTVNENQ NVDAKVKAEE SEIEKLTTKL
 101 ADTDAALADT DAALDATTNA LNKLGENITT FAEETKTNIV KIDEKLEAVA
 151 DTVDKHAEAF NDIADSLEDT NTKADEAVKT ANEAKQTAEE TKQNVDAKVK
 201 AAETAAGKAE AAAGTANTAA DKAEEAVAALKV TDIKADIATN KDNIAKKANS
 251 ADVYTREESD SKFVRIDGLN ATTEKLDTRL ASAEKSIADH DTRLNLGDKT
 301 VSDLRKETRQ GLAEQAALSG LFQPYNVGRF NVTAAVGGYK SESAVAIGTG
 351 FRFTENFAAK AGVAVGTSSG SSAAYHVGVN YEWGSGGGGV AADIGAGLAD
 401 ALTAPLDHKD KGLQSLTLDQ SVRKNEKLKL AAQGAEKTYG NGDSLNTGKL
 451 KNDKVSRFDF IRQIEVDGQL ITLESGEFQV YKQSHSALTA FQTEQIQDSE
 501 HSGKMKVAKRQ FRIGDIAGEH TSFDKLPEGG RATYRGTAFG SDDAGGKLTY
 15 551 TIDFAAKQGN GKIEHLKSPE LNVDLAAADI KPDGKRHAVI SGSVLYNQAE
 601 KGSYSLGIFG GKAQEYAGSA EVKTVNGIRH IGLAAKQLEH HHHHH*

961-983

20 1 ATGGCCACAA ACGACGACGA TGTAAAAAAA GCTGCCACTG TGGCCATTGC
 51 TGCTGCCTAC ACAATGCC AAGAAATCAA CGGTTCAAA GCTGGAGAGA
 101 CCATCTACGA CATTGATGAA GACGGCACAA TTACCAAAAA AGACGCAACT
 151 GCAGCCGATG TTGAAGCCGA CGACTTTAAA GGTCTGGTC TGAAAAAAAGT
 201 CGTACTAACG CTGACCAAAA CCGTCAATGA AAACAAACAA AACGTCGATG
 25 251 CCAAAGTAAA AGCTGCAGAA TCTGAAATAG AAAAGTTAAC AACCAAGTTA
 301 CGACAGACTG ATGCCGCTT AGCAGATACT GATGCCGCTC TGGATGCAAC
 351 CACCAACGCC TTGAATAAT TGGGAGAAAA TATAACGACA TTTGCTGAAG
 401 AGACTAAGAC AAATATCGTA AAAATTGATG AAAAATTAGA AGCCGTGGCT
 451 GATACCGTCG ACAAGCATGC CGAACGATTC AACGATATCG CCGATTTCATT
 50 501 GGATGAAACC AACACTAAGG CAGACGAAGC CGTCAAAACCC GCCAATGAAG
 551 CCAAACAGAC GGCGAAGAA ACCAAACAAA ACGTCGATGC CAAAGTAAA
 601 GCTGCAGAAA CTGAGCAGG CAAAGCCGAA GCTGCCGCTG GCACAGCTAA
 651 TACTGCAGCC GACAAGGGCG AAGCTGTCGC TGCAAAAGTT ACCGACATCA
 701 AAGCTGATAT CGCTACGAAC AAAGATAATA TTGCTAAAAAA AGCAAAACAGT
 751 GCCGACGTGT ACACCAAGAGA AGAGTCTGAC AGCAAAATTG TCAGAAATTGA
 801 TGGTCTGAAC GCTACTACCG AAAAATTGGA CACACGCTTG GCTTCITGCTG
 851 AAAAATCCAT TGCCGATCAC GATACTGCC TGAAACGGTT GGATAAAAACA
 901 GTGTCAGACC TGCGCAAAGA AACCCGCCAA GGCCTTGAG AACAAAGCCGC
 951 GCTCTCCGGT CTGTTCCAAC CTTACAACGT GGGTCGGTTC AATGTAACGG
 40 1001 CTGCAGTCGG CGGCTACAAA TCCGAATCGG CAGTCGCCAT CGGTACCGGC
 1051 TTCCGCTTTA CCGAAAACCT TGCCGCCAAA GCAGGGCTGG CAGTCGGCAC
 1101 TTCGTCCGGT TCTTCCGCCAG CCTACCATGT CGCGCTCAAT TACGAGTGGG
 1151 GATCCGGCGG AGGCGGCAG TCTGCCCGG ACTTCAATGC AGGCGGTACC
 1201 GGTATCGGCA GCAACAGCAG AGCAACAACA CGCAAATCAG CAGCAGTATC
 45 1251 TTACGCCGGT ATCAAGAACG AAATGTGAA AGACAGAACG ATGCTCTGTG
 1301 CCGGTGGGA TGACGTTGCG GTTACAGACA GGGATGCCAA AATCAATGCC
 1351 CCCCCCCCCGA ATCTGCATAC CGGAGACTTT CCAAACCCCAA ATGACCCATA
 1401 CAAGAACCTCA AACCTGCAAT TGAAGCAGGC TATACAGGAC
 1451 GCGGGGTAGA GGTAGGTATC TCGACACAG CGAATCCGT CGGCAGCATA
 50 1501 TCCCTTTCCCG AACTGTATGG CAGAAAAGAA CACGGCTATA ACGAAAATTA
 1551 CAAAAACTAT ACGGCGTATA TGCAGGAAGGA AGCGCCTGAA GACGGAGGCG
 1601 GTAAAGACAT TGAAGCTCTT TCGACGATG AGGCGCTTAT AGAGACTGAA
 1651 GCAAAGCCGA CGGATATCCG CCACGTTAAA GAAATCGGAC ACATCGATT
 55 1701 GGTCTCCCAT ATTATTGGCG GGCCTTCCGT GGACGGCAGA CCTGCAGGCG
 1751 GTATTGCGCC CGATGCGACG CTACACATAA TGAATACGAA TGATGAAACC
 1801 AAGAACGAAA TGATGGTGC AGCCATCCGC AATGCATGGG TCAAGCTGGG
 1851 CGAACGTGGC GTGCGCATCG TCAATAACAG TTTTGGAAACA ACATCGAGGG
 1901 CAGGCACTGC CGACCTTTTC CAAATAGCCA ATTCCGGAGGA GCAGTACCGC
 1951 CAAAGCGTTGC TCGACTATTC CGGGGGTGAT AAAACAGACG AGGGTATCCG
 60 2001 CCTGATGCAA CAGAGCGATT ACGGCAACCT GTCCCTACCA ATCCGTAATA
 2051 AAAACATGCT TTTCATCTT TCGACAGGCA ATGACGCACA AGCTCAGCCC
 2101 AACACATATG CCCTATTGCC ATTATATGAA AAAGACGCTC AAAAGGGCAT
 2151 TATCACAGTC GCAGGGCTAG ACCGGAGTGG AGAAAAGTTC AAACGGAAA
 2201 TGTATGGAGA ACCGGGTACA GAACCGCTTG AGTATGGCTC CAACCAATTGC
 2251 GGAATTACTG CCATGTGGT CCTGTCGGCA CCCTATGAAAG CAAGCGTCCG
 2301 TTTCACCCGT ACAAAACCCGA TTCAAATTGC CGGAACATCC TTTTCCGCAC
 2351 CCATCGTAAC CGGCACGGCG GCTCTGCTGC TGCAGAAATA CCCGTGGATG

2401	AGCAACGACA	ACCTGCGTAC	CACGTTGCTG	ACGACGGCTC	AGGACATCGG
2451	TGCAGTCGGC	GTGGACAGCA	AGTTCGGCTG	GGGACTGCTG	GATGCGGGTA
2501	AGGCCATGAA	CGGACCCCGG	TCCCTTCCGT	TCCGGCAGCTT	TACCGCCGAT
2551	ACGAAAGGTA	CATCCGATAT	TGCCCTACTCC	TTCCGTAACG	ACATTTCAGG
5 2601	CACGGGCGGC	CTGATCAAA	AAGGCAGCAG	CCAACATGCAA	CTGCACGGCA
2651	ACAACACCTA	TACGGGCAA	ACCATTATCG	AAGGCAGGTTTC	GCTGGTGTGTT
2701	TACGGCAACA	ACAATCGGA	TATGCGCGTC	GAAACCAAAG	GTGCGCTGAT
2751	TTATAACGGG	CGGGCATCCG	CGGGCAGCCT	GAACAGCGAC	GGCATTTGTCT
10 2801	ATCTGGCAGA	TACCGACCAA	TCCGGCGCAA	ACGAAACCGT	ACACATCAAA
2851	GGCAGTCTGC	AGCTGGACGG	CAAAGGTACG	CTGTACACAC	GTTTGGGCAA
2901	ACTGCTGAAA	GTGGACGGTA	CGGCAGATTAT	CGGCAGGCAAG	CTGTACATGT
2951	CGGCACGCGG	CAAGGGGCA	GGCTATCTCA	ACAGTACCGG	ACGACGTGTT
3001	CCCTTCCTGA	GTGCGGCCAA	AATCGGGCAG	GATTATTCTT	TCTTCACAAA
15 3051	CATCGAAACC	GACGGCGGCC	TGCTGGCTTC	CCTCGACAGC	GTCGAAAAAA
3101	CAGCGGGCAG	TGAAGGGCAG	ACGCTGTCCCT	ATTATGTCCG	TCCGGCGCAAT
3151	CGGGCACGGA	CTGCTTCGGC	AGCAGCACAT	TCCGGCGCCCG	CCGGTCTGAA
3201	ACACGCCGTA	AAACAGGGCG	GCAGCAATCT	GGAAAACCTG	ATGGTCAAC
3251	TGGATGCCCTC	CGAACATCAC	GCAACACCCG	AGACGGTTGA	AACTGCCGCA
20 3301	GCCGACCGCA	CAGATATGCC	GGGCATCCGC	CCCTACGGCG	CAACTTCCG
3351	CGCAGCGGCA	GCCGTACAGC	ATGCGAATGC	CGCCGACGGT	GTACGCATCT
3401	TCAACAGTCT	CGCCGCTACC	GTCTATGCCG	ACAGTACCGC	CGCCCATGCC
3451	GATATGCAGG	GACGCCGCT	GAAAGCCGTA	TCGGACGGGT	TGGACCAACAA
3501	CGGCACGGGT	CTGCGCGTCA	TCGCGCAAAC	CCAACAGGAC	GGTGGAACCGT
3551	GGGAAACAGGG	CGGTGTTGAA	GGCAAAATAGC	CGGGCAGTAC	CCAAACACTGTC
25 3601	GGCATTGCCG	CGAAAACCGG	CGAAAATACG	ACAGCAGCGG	CCACACTGGG
3651	CATGGGACGC	AGCACATGGA	GCGAAAACAG	TGCAAATGCA	AAAACCGACA
3701	GCATTAGTCT	GTGTTGCAGGC	ATACGGCAGC	ATGCGGGCGA	TATCGGCTAT
3751	CTCAAAGGCC	TGTTCTCTTA	CGGACGCTAC	AAAACAGCA	TCAGCCGCAG
3801	CACCGGTGCG	GACGAACATG	CGGAAGGCAG	CGTCAACGGC	ACGCTGATGC
30 3851	AGCTGGGCGC	ACTGGGCGGT	GTCAACGTT	CGTTTGCCGC	AACGGGAGAT
3901	TTGACGGTCG	AAGGCGGTCT	GCGCTACGAC	CTGCTCAAAC	AGGATGCATT
3951	CGCCGAAAAAA	GGCAGTGTCT	TGGGCTGGAG	CGGCAACAGC	CTCACTGAAG
4001	GCACGCTGGT	CGGACTCGCG	GGTCTGAAGC	TGTCGCAACC	CTTGAGCGAT
4051	AAAGCCGTC	TGTTTGCAAC	GGCGGGCGTG	GAACGCGAC	TGAACGGACG
35 4101	CGACTACACG	GTAAACGGCG	GCTTTACCGG	CGCGACTGCA	GCAACCGGCA
4151	AGACGGGGGC	ACGCAATATG	CCGCACACCC	GTCTGGTTGC	CGGCCTGGC
4201	CGGGATGTCG	AATTGGCAA	CGGCTGGAAC	GGCTTGGCAC	GTTACACGTA
4251	CGCCGGTTCC	AAACAGTACG	GCAACCCACAG	CGGACGAGTC	GGCGTAGGCT
4301	ACCGGTTCCCT	CGAGCACAC	CACCACCAAC	ACTGA	
40	1	MATNDDDVKK	AATVAIAAAY	NNGQEINGFK	AGETIYDIDE
	51	AADVEADDfk	GLGLKKVVNT	LTKTVNENKQ	NVDAKVAAE
	101	ADTDAALADT	DAALDATTNA	LNKLEGENIT	SEIEKLTTKL
	151	DTVDKHAEEAf	NDIADSIDET	NTKADEAVKT	FAEETKTNIV
	201	AAETAAGKAE	AAAGTANTAA	DKAEAVAAKV	KIDEKLEAVA
	251	ADVYTREESD	SKFVRIDGLN	ATTEKLDTRL	TDIKADIATN
	301	VSDLRKETRQ	GLAEQAALSG	LFQPYNVGRF	KDNIAKKANS
	351	FRFTENFAAK	AGVAVGTSSG	SSAAYHVGVN	AGGATGCATT
	401	GIGSNSRATT	AKSAAVSYAG	IKNEMCKDRS	YEWGSGGGGT
	451	PPPNLHTGDF	PNPNDAYKNL	MLCAGRDDVA	SAPDFNAGGT
	501	SFPELYGRKE	HGYNENYKNY	DTRLNGLDKT	VDTGESVGSI
	551	AKPTDIRHVK	EIGHIDLVSH	DTGRRGVEVGI	FDDEAVIETE
	601	KNEMMVAAIR	NAWVKLGERG	DGGGKDIEAS	LHIMNTNDET
	651	QALLDYSGGD	KTDEGIRLMO	PAGGIAPDAT	QIANSEEQYR
	701	NTYALLPFYE	KDAQKGITIV	TSRAGTADLF	STGNDQAQP
	751	GITAMWCLSA	PYEASVRFTR	QSDYGNLSYH	EPLEYGSNHC
	801	SNDNLRTTLL	TNAQDIDGAVG	IRKNKNMLFIF	GYLNSTGRRV
	851	TKGTSIDIAYS	FRNDISGTGG	LIKKGSQLQ	SFPFGDFTAD
	901	YGNNKSDMRV	ETKGALIYNG	LHGNNTYTGK	TIIEGGSLVL
	951	GSLQLDGKGT	IYTRLGKLK	GIVYLADTDQ	SGANETVHIK
	1001	PFLSAAKIGQ	DYSFFTNIET	SGANETVHIK	GYLNSTGRRV
	1051	AARTASAAA	DGGLLASLDS	VEKTAGSEGD	TLSYYVRRGN
	1101	ADRTDMPGIR	SAPAGLKHAV	MVELDASESS	ATPETVETAA
	1151	DMQGRRRLKAV	EQGGSNLENL	VRIIFNSLAAT	VYADSTAHA
	1201	GIAAKTGEENT	AVQHANAADG	GGTWEQGGVE	GKMRGSTQTV
	1251	LKGFLFSYGRY	TRVIAQTQOD	IRHDAGDIGH	VNPVFAATGD
	1301	LTVEGGLRYD	LLKQDAFAEK	GSALGWGSGNS	GLKLSQPLSD

1351 KAVLFATAGV ERDLNGRDYT VTGGFTGATA ATGKTGARMN PHTRLVAGLG
 1401 ADVEFGNGWN GLARYSYAGS KQYGNHSGRV GVGYRFLEHH HHHH*

5 **961c-ORF46.1**

1	ATGGCCACAA	ACGACGACGA	TGTTAAAAAA	GCTGCCACTG	TGGCCATTGC
51	TGCTGCCTAC	AAACATGCC	AAGAAATCAA	CGGTTCAAA	GCTGGAGAGA
101	CCATCTACGA	CATTGATGAA	GACGGCACAA	TTACCAAAAA	AGACGCAACT
151	GCAGCCGATG	TTGAAGCCGA	CGACTTTAAA	GGTCTGGGTC	TGAAAAAAAGT
201	CGTGACTAAC	CTGACCAAAA	CCGTCATGAA	AAACAAACAA	AACGTCGATG
251	CCAAAGTAAA	AGCTGCAGAA	TCTGAAATAG	AAAAGTAAAC	AACCAAGTTA
301	GCAGACACTG	ATGCGCTT	AGCAGATACT	GATGCCGCTC	TGGATGCAAC
351	CACCAACGCC	TTGAAATAAT	TGGGAGAAAA	TATAACGACA	TTTGCCTGAAG
401	AGACTAAGAC	AAATATCGTA	AAAAATTGATG	AAAATTAGA	AGCCGTGGCT
451	GATACCGTCG	ACAAGCATGC	CGAACATTC	ACGATATCG	CCGATTCAATT
501	GGATGAAACC	AAACACTAAGG	CAGACGAAGC	CGTCAAAACC	GCCAATGAAG
551	CCAAACAGAC	GGCCGAAGAA	ACCAAAACAA	ACGTCGATGC	CAAAGTAAAAA
601	GCTGCAGAAA	CTGCAGCAGG	CAAAGCCGA	GCTGCCGCTG	GCACAGCTAA
651	TACTGCAGCC	GACAAGGCCG	AAGCTGTCGC	TGAAAGAGTT	ACCGACATCA
701	AAGCTGATAT	CGCTACGAAC	AAAGATAATA	TTCGCTAAAAAA	AGCAAACAGT
751	GCCGACGTGT	ACACCAGAGA	AGAGTCTGAC	AGCAAATTG	TCAGAATTGA
801	TGGCTGAAAC	GCTACTACCG	AAAATTGGA	CACACGCTTG	GCTTCCTGCTG
851	AAAATCCAT	TGCCGATCAC	GATACTCGCC	TGAACGGTTT	GGATAAAAACA
901	GTGTCAGACC	TGCGCAAAGA	AACCCGCCAA	GGCCTTGCA	AACAAGCCGC
951	GCTCTCCGGT	CTGTTCCAAC	CTTACAACGT	GGGTGGATCC	GGAGGAGGAG
1001	GATCAGATTT	GGCAAACGAT	TCTTTTATCC	GGCAGGTTCT	CGACCGTCAG
1051	CATTCGAAC	CCGACGGAA	ATACCACCTA	TTCCGGCAGCA	GGGGGAACT
1101	TGCCGAGCGC	AGCGGCCATA	TCGGATTGGG	AAAATACAA	AGCCATCAGT
1151	TGGGCAACCT	GATGATTCAA	CAGGGGGCCA	TTAAAGGAAA	TATCGGCTAC
1201	ATTGTCGGCT	TTTCCGATCA	CGGGCACGAA	GTCCATTCCC	CCTTCGACAA
1251	CCATGCCCTCA	CATTCCGATT	CTGATGAAAGC	GGGTAGTCCC	GTTGACGGAT
1301	TTAGCCTTTA	CCGCATCCAT	TGGGACGGAT	ACGAACACCA	TCCCAGCGAC
1351	GGCTATGACG	GGCCACAGGG	GGCGGGCTAT	CCCGCTCCC	AAGGCGCGAG
1401	GGATATATAC	AGCTACGACA	TAAAAGCGT	TGCCCCAAAT	ATCCGCTCA
1451	ACCTGACCGA	CAACCGCAGC	ACCGGACAAC	GGCTTGCCGA	CCGTTTCCAC
1501	AATGCCGGA	GTATGCTGAC	GCAAGGAGTA	GGCGACGGAT	TCAAACGCGC
1551	CACCCGATAC	AGCCCCGAGC	TGGACAGATC	GGGCAATGCC	GCCGAAGCCT
1601	TCAACGGCAC	TGCAGATATC	GTTAAAAACAA	TCATCGGCGC	GGCAGGAGAA
1651	ATTGTCGGCG	CAGGCGATGC	CGTGCAGGGC	ATAAGCGAAG	GCTCAAACAT
1701	TGCTGTCATG	CACGGCTGG	GTCTGCTTTC	CACCGAAAAC	AAGATGGCGC
1751	GCATCAACGA	TTGGCAGAT	ATGGCGCAAC	TCAAAGACTA	TGCCGAGCA
1801	GCCATCCGCG	ATTGGCAGT	CCAAAACCCC	AATGCGCAC	AAGGCATAGA
1851	AGCCGTCAGC	AATATCTTA	TGGCAGCCAT	CCCCATCAA	GGGATGGAG
1901	CTGTTCGGGG	AAAATACGGC	TTGGGGCGCA	TCACGGCACA	TCCTATCAAG
1951	CGGTCGCGA	TGGGCGCGAT	CGCATTGCCG	AAAGGAAAT	CCGCGTCAG
2001	CGACAATTTC	CGCGATGCCG	CATACGCCAA	ATACCGTCC	CCTTACCAATT
2051	CCCGAAATAT	CCGTTCAAAC	TTGGAGCGC	GTTACGGCAA	AGAAAACATC
2101	ACCTCCTCAA	CCGTGCCGCC	GTCAAACGGC	AAAATGTCA	AACTGGCAGA
2151	CCAACGCCAC	CCGAAGACAG	GGCTACCGTT	TGACGGTAAA	GGGTTCCGA
2201	ATTTGAGAA	GCACGTGAAA	TATGATACGC	TCGAGCACCA	CCACCAACCAC
2251	CACTGA				
1	MATNDDDVKK	AATVAIAAA	NNGQEINGFK	AGETTYDIDE	DGTITKKDAT
51	AADVEADDK	GLGLKKVVTN	LTKTVNENKQ	NVDAKVAAE	SEIEKLTTKL
101	ADTDAALADT	DAALDATTNA	LNKLGENITT	FAEETKTNIV	KIDEKLEAVA
151	DTVDKHAEEAF	NDIADSLDET	NTKADEAVKT	ANEAKQTAEE	TKQNVDAKVK
201	AAETAAGKAE	AAAGTANTAA	DKAEAVAAKV	TDIKADIATN	KDNIAKKANS
251	ADVYTREESD	SKFVRIDGLN	ATTEKLDTRL	ASAEKSIADH	DTRLNGLDKT
301	VSDLRKETRQ	GLAEQAALSG	LFQPYNVGGS	GGGGSDLAND	SFIRQVLDRQ
351	HFEPDGKYHL	FGRSGELAER	SGHIGLGIQ	SHQLGNLMIQ	QAAIKGNIGY
401	IVRFSDHGHE	VHSPPFDNHAS	HSDSDEAGSP	VDGFSLYRIH	WDGYEHHPAD
451	GYDGPQGGGY	PAPKGARDIY	SYDIKGVQAQ	IRLNLTDRNS	TGQRLADRFH
501	NAGSMLTQGV	GDGFKRATRY	SPELDRSGNA	AEAFNGTADI	VKNIIIGAAGE
551	IVGAGDAVQG	ISEGNSIAVM	HGLGLLSTEM	KMARINDLAD	MAQLKDYAAA
601	AIRDWAVQNP	NAAQGIEAVS	NIFMAAIPIK	GIGAVRGKYG	LGGITAHPIK
651	RSQMGAIALP	KGKSAVSDNF	ADAAYAKYPS	PYHSRNIRSN	LEQRYGKENI
701	TSSTVPPSNG	KNVKLADQRH	PKTGVFDGK	GFPNFEKHVK	YDTLEHHHHH

751 H*

961c-741

5 1 ATGGCCACAA ACGACGACGA TGTTAAAAAA GCTGCCACTG TGGCCATTGC
 51 TGCTGCCTAC ACAATGGCC AAGAAATCAA CGGTTTCAA GCTGGAGAGA
 101 CCATCTACGA CATTGATGAA GACGGCACAA TTACCAAAAA AGACGCAACT
 151 GCAGCCGATG TTGAAGCCGA CGACTTTAAA GGTCTGGGTC TGAAAAAAAGT
 201 CGTGACTAAC CTGACCAAAA CCGTCAATGA AAACAAACAA AACGTCGATG
 251 CCAAAGTAAA AGCTGCAGAA TCTGAAATAG AAAAGTTAAC AACCAAGTTA
 301 GCAGACACTG ATGCCGCTT AGCAGATACT GATGCCGCTC TGGATGCAAC
 351 CACCAACGCC TTGAATAAT TGGGAGAAAA TATAACGACA TTTGCTGAAG
 401 AGACTAAGAC AAATATCGT AAAATTGATG AAAAATTAGA AGCCGTGGCT
 451 GATACCGTCG ACAAGCATGC CGAACGATT CCGATATCG CCGATTCATT
 501 GGATGAAACC AACACTAAGG CAGACGAAG CGTCAAAACC GCCAATGAAG
 551 CCAAACAGAC GGCGAAGAAA ACCAAACAA ACCTGATGC CAAAGAAAAA
 601 GCTGCAGAAA CTGCAGCAGG CAAAGCGAA GCTGCCGCTG GCACAGCTAA
 651 TACTGCAGCC GACAAGGCCG AAGCTGTCGC TGCAAAGTT ACCGACATCA
 701 AAGCTGATAT CGCTACGAAC AAAGATAATA TTGCTAAAAA AGCAAAACAGT
 751 GCCGACGTGT ACACCAGAGA AGAGTCGAC AGCAAATTG TCAGAATTGA
 801 TGGTCTGAAC GCTACTACCG AAAAATTGGA CACACGCTG GCTTCTGCTG
 851 AAAAATCCAT TGCCGATCAC GATACTCGCC TGAAACGGTT GGATAAAACA
 901 GTGTCAGACC TGCGCAAAAGA AACCCGCCAA GGCCCTTGCAG AACAAAGCCGC
 951 GCTCTCCGGT CTGTTCCAAC CTTACAAACGT GGGTGGATCC GGAGGGGGTG
 1001 GTGTCGCCGC CGACATCGGT GCGGGCTTG CGCATGCACT AACCGCACCG
 1051 CTCGACCATA AAGACAAAGG TTTGCACTG TTGACGCTGG ATCAGTCCGT
 1101 CAGGAAAAAC GAGAAACTGA AGCTGGCGGC ACAAGGTGCG GAAAAAAACTT
 1151 ATGGAAACGG TGACAGCCCTC AATAACGGCA AATTGAAGAA CGACAAGGTC
 1201 AGCCGTTTCG ACTTTATCCG CCAAATCGAA GTGGACGGGC AGCTCATTAC
 1251 CTTGGAGAGT GGAGAGTTCC AAGTATAACAA ACAAAAGCCAT TCCGCCTAA
 1301 CCGCCTTCA GACCGAGCAA ATACAAGATT CGGAGCATT CCGGAAGATG
 1351 GTTGCGAAAC GCCAGTTCAAG AATCGGCGAC ATAGCGGGCG AACATACATC
 1401 TTTTGACAAG CTTCCCGAAG GCGGCAGGGC GACATATCGC GGGACGGCGT
 1451 TCGGTTCAAG CGATGCCGGC GGAAAATCGA CCTACACCAT AGATTGCGCC
 1501 GCCAACCGAGG GAAACGGCAA AATCGAACAT TTGAAATCGC CAGAACTCAA
 1551 GTGTCACCTG GCCGCCGCCG ATATCAAGCC GGATGGAAA CGCCATGCCG
 1601 TCATCAGCGG TTCCGTCCTT TACAACCAAG CCGAGAAAGG CAGTTACTCC
 1651 CTCGGTATCT TTGGCGGAAA AGCCCAGGAA GTTGCCGGCA GCGCGGAAGT
 1701 GAAAACCGTA AACGGCATAAC GCCATATCGG CCTTGCCGCC AAGCAACTCG
 40 1751 AGCACCACCA CCACCACAC TGA

 1 MATNDDDVKK AATVAIAAAAY NNGQEINGFK AGETIYDIDE DGTITKKDAT
 51 AADVEADDFK GLGLKKVVTN LTKTVNENKQ NVDAKVKAEE SEIEKLTTKL
 101 ADTDAALADT DAALDATTNA LNKLGENITT FAEETKTNIV KIDEKLEAVA
 151 DTVDKHAEAF NDIADSLDET NTKADEAVKT ANEAKQTAEE TKQNVDAKVK
 201 AAETAAGKAE AAAGTANTAA DKAEEAVAAKV TDIKADIATN KDNIAKKANS
 251 ADVYTREESD SKFVRIDGLN ATTEKLDTRL ASAEKSIADH DTRLNGLDKT
 301 VSDLRKETRQ GLAEQAALSG LFQPYNVGGS GGGGVAADIG AGLADALTAP
 351 LDHKDKGLQS LTLDQSVRKN EKLKLAAQGA EKTYNGDSL NTGKLKNDKV
 401 SRFDFIRQIE VDGQLITLES GEFQVYKQSH SALTAFQTEQ IQDSEHSGKM
 451 VAKRQFRIGD IAGEHTSFDK LPEGGRATYR GTAEGSDDAG GKLTYTIDFA
 501 AKQGNKGIEH LKSPELNVDL AAADIKPDGK RHAVISGSVL YNQAEKGYS
 551 LGIFGGKAQE VAGSAEVKTV NGIRHIGLAA KQLEHHHHHH *

55

961c-983

60 1 ATGGCCACAA ACGACGACGA TGTTAAAAAA GCTGCCACTG TGGCCATTGC
 51 TGCTGCCTAC ACAATGGCC AAGAAATCAA CGGTTTCAA GCTGGAGAGA
 101 CCATCTACGA CATTGATGAA GACGGCACAA TTACCAAAAA AGACGCAACT
 151 GCAGCCGATG TTGAAGCCGA CGACTTTAAA GGTCTGGGTC TGAAAAAAAGT
 201 CGTGACTAAC CTGACCAAAA CCGTCAATGA AAACAAACAA AACGTCGATG
 251 CCAAACTAAA AGCTGCAGAA TCTGAAATAG AAAAGTTAAC AACCAAGTTA
 301 GCAGACACTG ATGCCGCTT AGCAGATACT GATGCCGCTC TGGATGCAAC
 351 CACCAACGCC TTGAATAAT TGGGAGAAAA TATAACGACA TTTGCTGAAG
 401 AGACTAAGAC AAATATCGT AAAATTGATG AAAAATTAGA AGCCGTGGCT
 451 GATACCGTCG ACAAGCATGC CGAACGATT AACGATATCG CCGATTTCATT
 501 GGATGAAACC AACACTAAGG CAGACGAAGC CGTCAAAACC GCCAATGAAG

551	CCAAACAGAC	GGCGAAGAA	ACCAAACAAA	ACGTCGATGC	CAAAGTAAAA
601	GCTGCAGAAA	CTGCAGCAGG	CAAAGCCGAA	GCTGCCGCTG	GCACAGCTAA
651	TACTGCAGCC	GACAAGGGCG	AAGCTGTCGC	TGCAAAAGTT	ACCGACATCA
701	AAGCTGATAT	CGCTACGAAC	AAAGATAATA	TTGCTAAAAAA	ACCAAACAGT
751	GCCGACGTGT	ACACCAGAGA	AGAGCTGAC	AGCAAATTG	TCAGAATTGA
801	TGGTCTGAAC	GCTACTACCG	AAAATTGGA	CACACGCTTG	GCTTCGCTG
851	AAAAAATCCAT	TGCCGATCAC	GATACTCGCC	TGAACGGTTT	CGATAAAAACA
901	GTGTCAGACC	TGCGCAAAGA	AACCCGCCAA	GGCCTTGCA	AACAAGCCGC
951	GCTCTCCGGT	CTGTTCCAAC	CTTACAACGT	GGGTGGATCC	GGCGGAGGCG
1001	GCACCTCTGC	GCCCCACTTC	AATGCAGGCG	GTACCGGTAT	CGGCAGCAAC
1051	AGCAGAGCAA	CAACAGCGAA	ATCAGCAGCA	GTATCTTACG	CCGGTATCAA
1101	GAACGAAATG	TGCAAAGACA	GAAGCATGCT	CTGTGCCGGT	CGGGATGACG
1151	TTGCGGGTAC	AGACAGGGAT	GCCAAAATCA	ATGCCCCCCC	CCCGAAATCTG
1201	CATACCGGAG	ACTTTCCAAA	CCCAAATGAC	GCATACAAGA	ATTGATCAA
1251	CCTCAAACCT	GCAATTGAAAG	CAGGCTATAC	AGGACGCGGG	GTAGAGGTAG
1301	GTATCGTCGA	CACAGGGCAA	TCCGTCGGCA	GCATATCCTT	TCCCGAACTG
1351	TATGGCAGAA	AAGAACACGG	CTATAACGAA	AATTACAAAA	ACTATACGGC
1401	GTATATGCGG	AAGGAAGCGC	CTGAAGACGG	AGGCGGTAAA	GACATTGAAG
1451	CTTCTTCGA	CGATGAGGCC	GTTATAGAGA	CTGAAGCAA	GGCGACGGAT
1501	ATCCGCCACG	AAAAAGAAAT	CGGACACATC	GATTTGGTCT	CCCATATTAT
1551	TGGCGGGCGT	TCCGTGGACG	GCAGACCTGC	AGGCGGTATT	GCGCCCGATG
1601	CGACGCTACA	CATAATGAAT	ACGAATGATG	AAACCAAGAA	CGAAATGATG
1651	GTGTCAGGCCA	TCCGCAATGC	ATGGGTCAAG	CTGGGCGAAC	GTGGCGTGC
1701	CATCGTCAAT	AACAGTTTG	GAACAACATC	GAGGGCAGGC	ACTGCCGACC
1751	TTTCCAAAT	AGCCAATTG	GAGGAGCAGT	ACCGCCAAGC	GTTGCTCGAC
1801	TATTCCGGCG	GTGATAAAAC	AGACCGAGGT	ATCCGCCGTA	TGCAACAGAG
1851	CGATTACGGC	AACTGTCT	ACCACATCCG	TAATAAAAAC	ATGCTTTCA
1901	TCTTTTCGAC	AGGCAATGAC	GCACAAGCTC	AGCCCAACAC	ATATGCCCTA
1951	TTGCCATT	ATGAAAAGA	CGCTCAAAAA	GGCATTATCA	CAGTCGCAGG
2001	CGTAGACCAC	AGTGGAGAAA	AGTCAACACG	GGAAATGTAT	GGAGAACCGG
2051	GTACAGAAC	GCTTGAGTAT	GGCTCCAACC	ATTGCGGAAT	TACTGCCATG
2101	TGGTGCCTGT	CGGCACCCCTA	TGAAGCAAGC	GTCCGTTTCA	CCCGTACAAA
2151	CCCGATTCAA	ATTGCCGAA	CATCCTTTTC	CGCACCCATC	GTAACCGGCA
2201	CGGGCGCTCT	GCTGTCGAG	AAATACCGT	GGATGAGCAA	CGACAACCTG
2251	CGTACCCACGT	TGCTGACGAC	GGCTCAGGAC	ATCGGTGCA	TCGGCGTGG
2301	CAGCAAGTTC	GGCTGGGAC	TGCTGGATGC	GGGTAAGGGC	ATGAACGGAC
2351	CCCGCTCCTT	TCCGTTCCGG	GACTTTACCG	CCGATACGAA	AGGTACATCC
2401	GATATTGCCT	ACTCCTTCCG	TAACGACATT	TCAGGCACGG	GGGGCCTGAT
2451	CAAAAAGGC	GGCAGCCAAAC	TGCAACTGCA	CGGCAACAAAC	ACCTATACGG
2501	GCAAAACCAT	TATCGAAGGC	GGTCGCTGG	TGTTGTACGG	CAACAAACAAA
2551	TCGGATATGC	GGCTGCCAAC	CAAAGGTGCG	CTGATTATA	ACGGGGCGGC
2601	ATCCGGCGGC	AGCCTGAACA	CGCACGGCAT	TGTCTATCTG	GCAGATACCG
2651	ACCAATCCGG	CGCAAACGAA	ACCGTACACA	TCAAAGGCAG	TCTGCACTG
2701	GACGGCAAAAG	GTACGCTGA	CACACGTTTG	GGCAAACATG	TGAAAGTGG
2751	CGGTACGGCG	ATTATCGGCG	GCAAGCTGTA	CATGTCGGCA	CGCGGCAAGG
2801	GGGCAGGCTA	TCTCAACAGT	ACCGGACGAC	GTGTTCCCTT	CCTGAGTGC
2851	GCCAAAATCG	GGCAGGATTA	TTCTTTCTTC	ACAAACATCG	AAACCGACGG
2901	CGGCCTGCTG	GCTTCCCTCG	ACAGCGTCGA	AAAAACAGCG	GGCAGTGAAG
2951	GCGACACGCT	GTCCTATTAT	GTCCGTCGCG	GCAATGCGC	ACGGACTGCT
3001	TCGGCAGCGG	CACATTCGC	GCCCCCGGT	CTGAAACACG	CCGTAGAACAA
3051	GGGCGGCAGC	AATCTGAA	ACCTGATGGT	CGAAGCTGGAT	GCCTCCGAAT
3101	CATCCGCAAC	ACCCGAGACG	GTTGAAACTG	CCGCAGCGGA	CCGCACAGAT
3151	ATGCGGGCA	TCCGCCCTA	CGGCCGA	TTCCGCGCAG	CCGCAGCCGT
3201	ACAGCATCG	AATGCCCG	ACGGTGTACG	CATCTTCAAC	AGTCTGCCG
3251	CTACCGTCTA	TGCGCAGACT	ACGCCGCC	ATGCCGATAT	GCAGGGACGC
3301	CGCCTGAAAG	CCGTATCGGA	CGGGTTGGAC	CACAACGGCA	GGGGTCTGCG
3351	CGTCATCGCG	CAAACCCAAAC	AGGACGGTGG	AACGTGGAA	CAGGGCGGTG
3401	TTGAAGGCAA	AATGCGCGC	AGTACCCAA	CGCTGGCAT	TGCCCGAA
3451	ACCGGCAGAA	ATACGACAGC	AGCCGCCACA	CTGGGCATGG	GACCCAGCAC
3501	ATGGAGCGAA	AACAGTGC	ATGCAAAAC	CGACAGCATT	AGTCTGTTG
3551	CAGGCATACG	GCACGATGCG	GGCGATATCG	GCTATCTAA	AGGCCGTGTT
3601	TCCCTACGGAC	GCTACAAAAA	CAGCATCAGC	CCGAGCACCG	GTGCGGACGA
3651	ACATGCGAA	GGCAGCGTC	ACGGCACGCT	GATGCACTG	GGCGCACTGG
3701	GCGGTGTCAA	CGTTCCGTTT	GCCGCAACGG	GAGATTGAC	GGTCCAAGGC
3751	GGTCTGCGCT	ACGACCTGCT	CAAACAGGAT	GCATTGCGCC	AAAAGGCAG
3801	TGCTTGGGC	TGGAGCGGCA	ACAGCCTCAC	TGAAGGCACG	CTGGTCGGAC
3851	TCGCGGGTCT	GAAGCTGCG	CAACCCTTGA	GCGATAAAAGC	CGTCCTGTTT

3901	GCAACGGCGG	GCGTGGAACG	CGACCTGAAC	GGACGCGACT	ACACGGTAAC	
3951	GGGCAGCTTT	ACCGGGCGGA	CTGCAGCAAC	CGGCAAGACG	GGGGCACCGA	
4001	ATATGCCGCA	CACCCGCTCTG	GTTGCCGGCC	TGGGCCGGGA	TGTCGAATT	
4051	GGCAACGGCT	GGAACGGCTT	GGCACGTTAC	AGCTACGCCG	GTTCCAAACA	
5	4101	GTACGGCAAC	CACAGCGGAC	GAGTCGGCGT	AGGCTACCGG	TTCCTCGAGC
	4151	ACCACCAACCA	CCACCACTGA			
10	1	MATNDDDVKK	AATVAAIAAY	NNGQEINGFK	AGETIYDIDE	DGTITKKDAT
	51	AADVEADDK	GLGLKKVVTN	LTKTVNENQ	NVDAKVAAE	SEIEKLTTL
	101	ADTDAALADT	DAALDAATTNA	LNKLEGENITT	FAEETKTNIV	KIDEKLEAVA
	151	DTVDKHAEEAF	NDIADSLDET	NTKADEAVKT	ANEAKQTAEE	TKQNVDAKV
	201	AAETAAGKAE	AAAGTANTAA	DKAEAVAAKV	TDIKADIATN	KDNIAKKANS
	251	ADVYTREESD	SKFVRIDGLN	ATTEKLDTTRL	ASAEKSIADH	DTRLNGLDKT
15	301	VSDLRKETRQ	GLAEQAALSG	LFQPYNVGGS	GGGGTSAPDF	NAGGTGIGSN
	351	SRATTAKSAA	VSYAGIKNEM	CKDRSMLCAG	RDDVAVTDRD	AKINAPPNL
	401	HTGDFPNPND	AYKNLNLKPN	AIEAGYTGRG	VEVGIVDTGE	SVGSISFPEL
	451	YGRKEHGYNE	NYKNYTAYMR	KEAPEDGGGK	DIEASFDDEA	VIETEAKPTD
	501	IRHVKEIGHI	DLVSHIIGGR	SVDGRPAGGI	APDATLHIMN	TNDETKNEMM
	551	VAAIRNAWVK	LGERGVRIVN	NSFGTTSRAG	TADLFQIANS	EEQYRQALLD
20	601	YSGGDKTDEG	IRLMQQSDFY	NLSYHIRNKN	MLFIFSTGND	AQAQPNTYAL
	651	LFFYEKDAQK	GIITVAGVDR	SGEKFKREMY	GEPEGTTEPLEY	GSNHCGITAM
	701	WCLSAPYEAS	VRFTRTNPPIQ	IACTTSFSAPI	VTGTAALLLQ	KYPWMSNDNL
	751	RTTLLTTAQD	IGAVGVDSKF	GWGLLDAGKA	MNGPASFPFG	DFTADTKGTS
25	801	DIAYSFRNDI	SGTGGLIKKG	GSQLQLHGNN	TYTGKTIIEG	GSLVLYGNNK
	851	SDMRVETKGA	LIYNGAASGG	SLNSDGIVYL	ADTDQSGANE	TVHIKGSQL
	901	DGKGTLYTRL	GKLLKVDGTA	IIGGKLYMSA	RGKGAGYLN	TGRRVPFLSA
	951	AKIGQDYSFF	TNIETDGLL	ASLDSVEKTA	GSEGDTLSYY	VRRGNAARTA
	1001	SAAAHSAAPAG	LKHAVEQGGS	NLENLVMVELD	ASESSATPET	VETAAADRTD
	1051	MPGIRPYGAT	FRAAAAVQHA	NAADGVRIFN	SLAATVYADS	TAAHADMQGR
30	1101	RLKAVSDGLD	HNGTGLRVIA	QTQQDGGTWE	QGGVEGKMRG	STQTVGIAAK
	1151	TGENITAAAT	LGMGRSTWSE	NSANAKTDSI	SLFAGIRHDA	GDIGYIKGLF
	1201	SYGRYKNSIS	RSTGADEHAE	GSVNGTLMQL	GALGGGVNVPF	AATGDLTVEG
	1251	GLRYDPLLQD	AFAEKGSALG	WSGNSLTEGT	LVGLAGLKL	QPLSDKAVLF
	1301	ATAGVERDLN	GRDYTVTGGF	TGATAATGKT	GARNMPHTRL	VAGLGADVEF
35	1351	GNGWNGLARY	SYAGSKQYGN	HSGRVGVGYR	FLEHHHHHH*	

961cL-ORF46.1

40	1	ATGAAACACT	TTCCATCCAA	AGTACTGACC	ACAGCCATCC	TTGCCACTTT
	51	CTGTAGCGGC	GCACTGGCAG	CCACAAACGA	CGACGATGTT	AAAAAAGCTG
	101	CCACTGTGGC	CATTGCTGCT	GCCTACAAACA	ATGGCCAAGA	AATCAACCGT
	151	TTCAAAGCTG	GAGAGACCAT	CTACGACATT	GATGAAGACG	GCACAATTAC
	201	CAAAAAAGAC	GCAACTGCAG	CCGATGTTGA	AGCCGACGAC	TTTAAAGGTC
	251	TGGGTCTGAA	AAAAGTCGTG	ACTAACCTGA	CCAAAACCGT	CAATGAAAAC
	301	AAACAAAACG	TCGATGCCAA	AGTAAAAGCT	GCAGAATCTG	AAATAGAAAA
	351	GTAAACAACC	AAGTTAGCAG	ACACTGATGC	CGCTTTAGCA	GATACTGATG
	401	CCGCTCTGGA	TGCAACCACC	AACGCCCTGA	ATAAATTGGG	AGAAAATATA
	451	ACGACATTTG	CTGAAGAGAC	TAAGACAAAT	ATCGTAAAAA	TTGATGAAAA
	501	ATTAGAAGCC	GTGGCTGATA	CCGTCGACAA	GCATGCCGAA	GCATTCAACG
	551	ATATCGCCG	TTCATTGGAT	GAAACCAACA	CTAAGGCAGA	CGAAGCCGTC
	601	AAAACCGCCA	ATGAAGCCAA	ACAGACGGCC	GAAGAACCA	AACAAAACGT
	651	CGATGCCAA	GTAAAAGCTG	CGAAAACTGC	AGCAGGCAA	GCGGAAGCTG
	701	CCGCTGGCAC	AGCTAATACT	GCAGCCGACA	AGGCCGAAAGC	TGTCGCTGCA
	751	AAAGTTACCG	ACATCAAAGC	TGATATCGCT	ACGAACAAAG	ATAAATTATTGC
	801	TAAAAAAAGCA	AACAGTGCCG	ACGTGTACAC	CAGAGAAGAG	TCTGACAGCA
	851	AATTTGTCA	AATTGATGGT	CTGAACGCTA	CTACCGAAAA	ATTGGACACA
	901	CGCTTGGCTT	CTGCTAAAAA	ATCCATTGCC	GATCACGATA	CTCGCCTGAA
	951	CGGTTTGGAT	AAAACAGTGT	CAGACCTGCG	CAAAGAACCC	CGCCAAGGCC
60	1001	TTGCAGAACAA	AGCCGCGCTC	TCCGGTCTGT	TCCAACCTTA	CAACGTGGGT
	1051	GGATCCGGAG	GAGGAGGATC	AGATTTGGCA	AACGATTCTT	TTATCCGGCA
	1101	GGTTCTCGAC	CGTCAGCATT	TCGAACCCGA	CACCTATT	CGC
	1151	GCAGCAGGGGG	GGAACTTGCC	GAGCGCAGCG	GCCATATCGG	ATTGGAAAAA
	1201	ATACAAAGCC	ATCAGTTGGG	CAACCTGTATG	ATTCAACAGG	CGGCCATTAA
	1251	AGGAAATATC	GGCTACATTG	TCCGCTTTTC	CGATCACGGG	CACGAAGTCC
	1301	ATTCCCCCTT	CGACAACCAT	GCCTCACATT	CCGATTCTGA	TGAAGCCGGT
	1351	AGTCCCCTTG	ACGGATTAG	CCTTACCGC	ATCCATTGGG	ACGGATACGA
65	1401	ACACCACCCC	GCGGACGGCT	ATGACGGGCC	ACAGGGCGC	GGCTATCCCG

1451	CTCCCAAAGG	CGCGAGGGAT	ATATACAGCT	ACGACATAAA	AGGC GTT GCC	
1501	CAAAATATCC	GCCTCAACCT	GACCGACAAC	CGCAGCACCG	GACAACGGCT	
1551	TGCCGACCGT	TTCCACAATG	CCGGTAGTAT	GCTGACGCAA	GGAGTAGGCG	
1601	ACGGATTCAA	ACCGCGCAC	CGATACAGCC	CCGAGCTGG	CAGATCGGGC	
5	1651	AATGCCGCG	AAGCCTTCAA	CGGC ACTGCA	GATATCGTTA	AAAACATCAT
1701	CGGC GCGGCA	GGAGAAATTG	TCGGCGCAGG	CGATGCCGTG	CAGGGCATAA	
1751	GCGAAGGCTC	AAACATTGCT	GTCATGCACG	GCTTGGGTCT	GCTTCCACC	
1801	GAAAACAAGA	TGGCGCGCAT	CAACGATTG	GCAGATATGG	CGCAACTCAA	
10	1851	AGACTATGCC	GCAGCAGGCC	TCCCGGATTG	GGCAGTCCAA	AACCCCAATG
1901	CCGCACAAGG	CATAGAAGCC	GTCAGCAATA	TCTTTATGGC	AGCCATCCCC	
1951	ATCAAAGGGA	TTGGAGCTGT	TCGGGGAAAA	TACGGCTTGG	GC GG CATCAC	
2001	GGCACATCCT	ATCAAGCGGT	CGCAGATGGG	CGCGATCGCA	TTGCCGAAAG	
2051	GGAAATCCGC	CGTCAGCGAC	AATTTCGCG	ATGCCGGCATA	CGCCAAATAC	
15	2101	CCGTCCCCCT	ACCATTCCCG	AAATATCCGT	TCAAAACTTGG	ACCAGCGTTA
2151	CGGAAAGAA	AAACATCACCT	CCTCAACCGT	GGCCGCGTCA	AACGGCAAAA	
2201	ATGTCAAAC	GGCAGACCAA	CGCCACCCG	AGACAGCGT	ACCGTTTGAC	
2251	GGTAAAGGGT	TTCCGAATT	TGAGAACAC	GTGAAATATG	ATACGTAAC	
2301	CGAG					
20	1	MKHFPSKVLT	TAILATFCSG	ALAATNDDDV	KKAATVAIAA	AYNNGQEING
	51	FKAGETIYDI	DEDGTITKKD	ATAADVEADD	FKGLGLKKVV	TNLTKTVNEN
	101	KQNVDAKVKA	AESEIEKLT	KLADTDAALA	DTDAALDATT	NALNKLGENI
	151	TTFAEETKTN	IVKIDEKLEA	VADTVDKHAE	AFNDIADSLD	ETNTKADEAV
25	201	KTANEAKQTA	EETKQNVDAK	VKAETAAGK	AEAAAAGTANT	AADKAEAVAA
	251	KVTDIKADIA	TNKDNIAKKA	NSADVYTREE	SDSKFVRIDG	LNATTEKLDT
	301	RLASAEKSIA	DHDTRLNGLD	KTVSDLRKET	RQGLAEQAA	SGLFQPYNVG
	351	GSGGGGSDLA	NDSFIRQVLD	RQHFEPDGKY	HLFGSRGELA	ERSGHIGL GK
	401	IQSHQLGNLM	IQQAAIKGNI	GYIVRFSDHG	HEVHSPFDNH	ASHSDSDEAG
30	451	SPVDGFSLYR	IHWGDYEHHP	ADGYDGPQGG	GYPAPKGARD	IYSYDIKGVA
	501	QNIRNLNLTDN	RSTGQRLADR	FHNAGSMLTQ	GVGDGFKRAT	RYSPELDRSG
	551	NAAEAFNGTA	DIVKNIIGAA	GEIVGAGDAV	QGISEGSNIA	VMHGLGLLST
	601	ENKMARINDL	ADMAQLKDYA	AAAIRDWAVQ	NPNAAQGIEA	VSNIFMMAIP
	651	IKGIGAVRKG	YGLGGITAHP	IKRSQMGATA	LPKGKSAVSD	NFADAAYAKY
35	701	PSPYHSRNIR	SNLEQRYGE	NITSSTVPPS	NGKNVKLADQ	RHPKTGVPFD
	751	GKGFPNFEKH	VKYDT*			

961cL-741

40	1	ATGAAACACT	TTCCATCCAA	AGTACTGACC	ACAGCCATCC	TTGCCACTTT
	51	CTGTAGCGGC	GCACTGGCAG	CCACAAACGA	CGACGATGTT	AAAAAAGCTG
	101	CCACTGTGGC	CATTGCTGCT	GCCTACAACA	ATGGCCAAGA	AATCAACGGT
	151	TTCAAAGCTG	GAGAGACCAT	CTACGACATT	GATGAAGACG	GCACAATTAC
	201	CAAAAAGAC	GCAACTGCG	CCGATGTTGA	AGCCGACGAC	TTTAAAGGTC
45	251	TGGGTCTGAA	AAAAGTCGTG	ACTAACCTG	CCAAAACCGT	CAATGAAAAC
	301	AAACAAAACG	TCGATGCCA	AGTAAAAGCT	CGAGAACATCTG	AAATAGAAAA
	351	GTTAACAAAC	AAGTTAGCAG	ACACTGATGC	CGCTTTAGCA	GATACTGATG
	401	CCGCTCTGGA	TGCAACCACC	AACGCCCTGA	ATAAAATTGGG	AGAAAATATA
	451	ACGACATTG	CTGAAGAGAC	TAAGACAAAT	ATCGTAAAAAA	TTGATGAAAA
50	501	ATTAGAACCC	GTGGCTGATA	CGCTGACAA	GCATGCCGAA	GCATTCAACG
	551	ATATCGCCGA	TTCATTGGAT	GAAACCAACA	CTAAGGCAGA	CGAAGCCGTC
	601	AAAACCGCCA	ATGAAGCCAA	ACAGACGGCC	GAAGAAACCA	AACAAAACGT
	651	CGATGCCAAA	GTAAAAGCTG	CAGAAACTGC	AGCAGGCCAA	GCCGAAGCTG
	701	CCGCTGGCAC	AGCTAATACT	GCAGCCGACAA	AGGCCGAAGC	TGTCGCTGCA
55	751	AAAGTTACCG	ACATCAAAGC	TGATATCGCT	ACGAACAAAG	ATAATATTGC
	801	AAAAAAAGCA	AACAGTGCCG	ACGTGTACAC	CAGAGAAAGAG	TCTGACAGCA
	851	AATTGATGGT	CTGAACGCTA	CTACCGAAAAA	ATTGGACACA	
	901	CGCTTGGCTT	CTGCTAAAAA	ATCCATTGCC	GATCACGATA	CTCGCCTGAA
	951	CGGTTTGGAT	AAAACAGTGT	CAGACCTGCG	CAAAGAAACC	CGCCAAGGCC
60	1001	TTGCAGAACAA	AGCCGCGCTC	TCCGGTCTGT	TCCAACCTTA	CAACGTGGGT
	1051	GGATCCGGAG	GGGGTGGTGT	CGCCGCCGAC	ATCGGTGCGG	GGCTTGCCGA
	1101	TGCACTAAC	GCACCGCTCG	ACCATAAAGA	CAAAGGTGTTG	CAGTCCTTGA
	1151	CGCTGGATCA	GTCCGTCAAG	AAAAACGAGA	AACTGAAGCT	GGCGGCACAA
	1201	GGTGCAGAAA	AAACTTATGG	AAACGGTGAC	AGCCTCAATA	CGGGCAAATT
65	1251	GAAGAACGAC	AAGTCAGCC	TTTTCGACTT	TATCCGCCAA	ATCGAAGTGG
	1301	ACGGGGCAGCT	CATTACCTG	GAGAGTGGAG	AGTTCCAAGT	ATACAAACAA
	1351	AGCCATTCCG	CCTTAACCGC	CTTTCAGACCC	GACCAAATAC	AAGATTGG
	1401	GCATTCCGGG	AAGATGGTTG	CGAAACGCCA	GTTCAGAAC	GGCGACATAG

5	1451	CGGGCGAAC	TACATCTTT	GACAAGCTTC	CCGAAGGC	GG CAGGGCGACA
	1501	TATCGCGGG	CGGCGTT	CGG	TTCAGACGAT	GCCGGCGGAA AACTGACCTA
	1551	CACCATAGAT	TTCGCCGCCA	AGCAGGGAAA	CGGCAA	AAATC GAACATTG
	1601	AATCGCCAGA	ACTCAATGTC	GACCTGGCCG	CGCCG	GATAT CAAGCCGGAT
	1651	GGAAAACGCC	ATGCGTCAT	CAGCGGT	TC	TACCA ACCAAGCCGA
	1701	GAAAGGCAGT	TACTCCCCTCG	GTATCTTGG	CGGAAAAGCC	CAGGAAGTTG
	1751	CCGGCAGCGC	GGAAGTGAAA	ACCGTAAACG	GCATACGCCA	TATCGCC
	1801	GCCGCCAAC	AACTCGAGCA	CCACCA	CCAC	CACTGA
10	1	MKHFPSKVLT	TAILATFC	SG ALAATN	DDDV KKAATV	IAA AYNNGQEING
	51	FKAGETIYDI	DEDGT	ITKKD ATAAD	VEADD FKGLGL	KVV TNLT
	101	KQNVDAKVKA	AESEIE	KLTT KLA	D TDAAL	DATT NALN
	151	TTFAEETKTN	IVKIDEKLEA	VADTV	DKHAE AFNDIAD	SLD ETNTK
	201	KTANEAKQTA	EETKQNVDAK	VKA	AAETAAGK AAAAAAGT	TANT AADKAE
	251	KVTDIKADIA	TNKDI	AKKA NSAD	VYTREE SDSKFVR	IDG LNATTE
	301	RLASAESKIA	DHDTRLNGLD	KTV	SDLRKET RQGLAEQ	AAAL SGLFQPV
	351	GSGGGGVAAD	IGAGLADALT	APLDHKDKGL	QSLTLDQSVR	KNEKL
	401	GAEKTYGN	SLNTGKLKD	KVSRDF	FIRO IEVDGQLITL	ESGEFQVYKQ
	451	SHSALTAFQ	EQIQDSEHSG	KMVA	KRQFRI GDIAGEHT	TSF DKLPEGGR
	501	YRGTAFGSDD	AGGKL	TYTID FAAKQ	QNGKI EHLK	SPELNV DLAAADIK
	551	GKRHAVISGS	VLYNQAEKGS	YSLGIF	GGKA QEVAG	SAEVK TVNGIRH
	601	AAKQLEHHHH	HH*			
25	<u>961cL-983</u>					
	1	ATGAAAAC	ACT TTCC	CATCAA AGTACT	GACC ACAG	CCATCC TTGCC
	51	CTGTAGCG	GCAG	CTGGCAG CCAC	AAACGA CGAC	GATGTT AAAAAGCTG
	101	CCACTGT	GGC CATT	GTGCTGC	GCCTACAA	ACGATGTT
	151	TTCAAAGCTG	GAGAGAC	CATT GATGAC	CTACGAC	ATT GAGAC
	201	CAAAAAGAC	GCAACT	GCAG	CGAC	TCAG
	251	TGGGTCTGAA	AAAAGTC	GTG ACTAAC	CTGA CCAAG	CTG CCGT
	301	AAACAAAACG	TCGAT	ATG CCAA AGT	GCAGAAT	CTG AAATAG
	351	GTAAACAAAC	AAGT	AGTAGCAG	ACACTGAT	GC CGCTT
	401	CCGCTCTGGA	TGCAAC	ACC AACG	CTGTA	CTG AGA
	451	ACGACATTG	CTGAGAGAC	AAAGCAAAT	ATCGT	AAAAAAT
	501	ATTAGAAGCC	GTGGC	TGATA	CGAC	CTAAG
	551	ATATCGCCG	TTCATTG	GGAT	ACAA	CTAAGG
	601	AAAACGCCA	ATGAAG	CCAA	ACAGAC	GGCC
	651	CGATGCCAA	GTAAAAG	CTG	CAGAA	ACTGC
	701	CCGCTGGC	ACGTA	AACT	ACTGC	AGCAGG
	751	AAAAGTTACCG	ACAT	CAAAG	CGAC	TGTCG
	801	TA	AAAGTC	AGT	ACAC	CGCA
	851	AATTGTCAG	AATTG	ATG	GTAC	CGCT
	901	CGCTTGGC	CTG	CTGAA	AC	CTG
	951	CGGTTGGAT	AAAACAG	TGT	CAGAC	TGCG
	1001	TTGCAGAAC	AGCCG	GCTC	TCCGG	TCTGT
	1051	GGATCCGGCG	GAGGCG	GGC	GGCC	GGCT
	1101	CGGTATCGC	AGCAAC	AGCA	AGC	AGCT
	1151	CTTACGCCG	TATCAAG	AA	AGAC	AGAG
	1201	GCCGGT	CGGG	ATGAC	GTG	GGGCC
	1251	CCCCCCCC	ATG	CGT	GGT	GGGAT
	1301	ACAAGAATT	CTG	CAGAC	TACAG	CCCA
	1351	CGGGGGTAG	AGGT	AGGT	ACAG	GGACT
	1401	ATCCTTCCC	GAAC	TGTATG	GCAG	AAAGA
	1451	ACAAAAC	TACGG	GTATG	GGGAG	ACAGC
	1501	GGTAAAGACA	TTGAAG	CTTC	GGAC	GGCTT
	1551	AGCAAAGCCG	ACGG	GGAC	GGG	GGCTT
	1601	TGGTCT	CC	ATTATTGGC	GGGCG	GGCG
	1651	GGTATT	GGC	GGC	GGT	GGC
	1701	CAAGAACGAA	ATG	ATG	GGT	GGC
	1751	CGGAACGTG	CGT	CGC	GGT	GGC
	1801	GCAGGC	ACTG	GTCA	AA	CGAG
	1851	CCAAGCG	CTCG	AA	TT	GGAGG
	1901	GCCTGATGCA	ACAGAG	GGAT	TACGG	CA
	1951	AAAAACATGC	TTT	CTG	GGAC	GGC
	2001	CAACACATAT	GCC	TATG	GGAC	GGC
	2051	TTATCACAGT	CGCAG	GGT	GGAG	GGAA

2101	ATGTATGGAG	AACCGGGTAC	AGAACCGCTT	GAGTATGGCT	CCAACCATTG	
2151	CGGAATTACT	GCCATGTGGT	GCCTGTCGGC	ACCCATGAA	GCAAGCGTCC	
2201	GTTCACCCG	TACAAACCCG	ATTCAAATTG	CGGAAACATC	CTTTTCCGCA	
2251	CCCATCGAA	CCGGCACCGC	GGCTCTGCTG	CTGCAGAAAT	ACCCGTGGAT	
5	2301	GAGCAACGAC	AACCTGCGTA	CCACGTTGCT	GACGACGGCT	CAGGACATCG
2351	GTGCAGTCGG	CGTGGACAGC	AAAGTCGGCT	GGGGACTGCT	GGATGCGGGT	
2401	AAGGCCATGA	ACGGACCCGC	GTCCCTTCCG	TTCGGCGACT	TTACCGCCGA	
2451	TACGAAAGGT	ACATCCGATA	TTGCTTACTC	CTTCCGTAAC	GACATTTCA	
10	2501	GCACGGGCGG	CCTGATCAA	AAAGGCGGCA	GCCAACTGCA	ACTGCACGGC
2551	AACAACACCT	ATACGGGCAA	AACCATTATC	GAAGGCGGTT	CGCTGGTGT	
2601	GTACGGCAAC	AACAAATCGG	ATATGCGCGT	CGAAACCAA	GGTGCGCTGA	
2651	TTTATAAACGG	GGCGGCATCC	GGCAGCAGCC	TGAACAGCGA	CGGCATTGTC	
15	2701	TATCTGGCAG	ATACCGACCA	ATCCGGCGCA	AACGAAACCG	TACACATCAA
2751	AGGCAGTCCTG	CAGCTGGACG	GCAGGAGGTAC	GCTGTACACA	CGTTGGGCA	
2801	AACTGCTGAA	AGTGGACGGT	ACGGCGATT	TCGGCGGCAA	GCTGTACATG	
2851	TCGGCACGCG	GCAAGGGGGC	AGGCTATCTC	AACAGTACCG	GACGACGTGT	
2901	TCCCTTCTG	AGTGCCTGCCA	AAATCGGGCA	GGATTATTCT	TTCTTCACAA	
2951	ACATCGAAAC	CGACGGCGGC	CTGCTGGCTT	CCCTCGACAG	CGTCGAAAAAA	
20	3001	ACAGGGGCA	GTGAAGGGCA	CACGCTGTCC	TATTATGTCC	GTCGCGGCAA
3051	TGCGGCACGG	ACTGCTTCGG	CAGCGGCACA	TTCCCGGCC	GCCGGTCTGA	
3101	AACACGCCGT	AGAACAGGGC	GGCAGCAATC	TGGAAAACCT	GATGGTCGAA	
3151	CTGGATGCCT	CCGAATCATC	CGCAACACCC	GAGACGGTTG	AAACTGCGGC	
3201	AGCCGACCGC	ACAGATATGC	CGGGCATCCG	CCCCTACGCC	GCAACTTCC	
3251	GGCAGCGGCC	AGCCGTACAG	CATGCGAATG	CCGGCGACGG	TGTACGCATC	
25	3301	TTCAACAGTC	TCGCGCTAC	CGTCTATGCC	GACAGTACCG	CCGCCCATGC
3351	CGATATGCAG	GGACGCCGCC	TGAAAGCCGT	ATCGGACGGG	TTGGACCACA	
3401	ACGGCACGGG	TCTGCGCGTC	ATCGCGCAA	CCCAACAGGA	CGGTGGAACG	
3451	TGGGAACAGG	GCGGTGTGA	AGGCAAAATG	CGCGGCAGTA	CCCAAACCGT	
3501	CGGCATTGCC	GCGAAAACCG	GCGAAAATAC	GACAGCAGCC	GCCACACTGG	
30	3551	GCATGGGACG	CAGCACATGG	AGCGAAAACA	GTGCAAATGC	AAAAACCGAC
3601	AGCATTAGTC	TGTTTGCAGG	CATACGGCAC	GATGCGGGCG	ATATCGGCTA	
3651	TCTCAAAAGG	CTGTTCTCT	ACGGACGCTA	AAAAAACAGC	ATCAGCCGCA	
3701	GCACCGGTGC	GGACGAACAT	GCGGAAGGCC	GCGTCAACGG	CACGCTGATG	
3751	CAGCTGGCG	CACTGGCGG	TGTCAACGTT	CCGTTTGCCG	CAACGGGAGA	
35	3801	TTTGACGGTC	GAAGGGCGTC	TGCGCTACGA	CCTGCTCAA	CAGGATGCAT
3851	TCGCCGAAAA	AGGCAGTGCT	TTGGGCTGG	CGGCAACAG	CCTCACTGAA	
3901	GGCACGCTGG	TCGGACTCGC	GGGTCTGAAG	CTGTCGCAAC	CCTTGAGCGA	
3951	TAAAGCCGTC	CTGTTGCCA	CGGGGGCGT	GGAACGCGAC	CTGAACGGAC	
40	4001	GCGACTACAC	GGTAACGGGC	GGCTTACCG	GCGCGACTGC	AGCAACCGGC
4051	AAGACGGGGG	CACGCAATAT	GCCGCACACC	CGTCTGGTTG	CCGGCCTGGG	
4101	CGCGGATGTC	GAATTGCGCA	ACGGCTGGAA	CGGCTTGGCA	CGTTACAGCT	
4151	ACGCCGGTTC	CAAACAGTAC	GGCAACCACA	GCGGACGAGT	CGGCGTAGGC	
4201	TACCGGTTCT	GACTCGAG				
45	1	MKHFPSKVLT	TAILATFCSG	ALAATNDDDV	KKAAATVAIAA	AYNNQQEING
51	51	FKAGETIYDI	DEDGTITKKD	ATAADVEADD	FKGLGLKKVV	TNLTKTVNEN
101	101	KQNVDAVKVA	AESEIEKLTT	KLADTDAAALA	DTDAALDATT	NALNKLGENI
151	151	TTFAEETKTN	IVKIDEKLEA	VADTVDKHAE	AFNDIADSID	ETNTKADEAV
201	201	KTANEAKQTA	EETKQNVDAK	VKAAETAAGK	AEAAAAGTANT	AADKAEAVAA
251	251	KVTDIKADIA	TNKDNIAKKA	NSADVYTREE	SDSKFVRIDL	LNATTEKLD
301	301	RLASAESKIA	DHDTRLNGLD	KTVSDLRKET	RQGLAEQAA	SGLFQPYNVG
351	351	GSGGGGTSAP	DFNAGGTGIG	SNSRATTAKS	AAVSYAGIKN	EMCKDRSMLC
401	401	AGRDDVAVTD	RDAKINAPPP	NLHTGDFFPNP	NDAYKNLNL	KPAIEAGYTG
451	451	RGVEVGIVDT	GESVGSISFP	ELYGRKEHGY	HIDLVSHIIG	GRSVDGRPAG
50	501	GKDIIEASFDD	EAVIETEAKP	TDIRHVKEIG	HIDLVSHIIG	GRSVDGRPAG
551	551	GIAPDATLHI	MNTNDETKNE	MMVAIRNAW	VKLGERGVRI	VNNSFGTTSR
601	601	AGTADLFBQIA	NSEEQYRQAL	LDYSGGDKTD	EGIRLMQQSD	YGNLSYHIRN
651	651	KNMLFIFSTG	NDAQAQPNTY	ALLPFYEKDA	QKGIIITVAGV	DRSGEKFKRE
701	701	MYGEPGTEPL	EYGSNHCGIT	AMWCLSAPYE	ASVRFTRTNP	IQIAGTSFSA
751	751	PIVTGTAALL	LQKYPWMSND	NLRTTLLTTA	QDIGHAVGVD	KFGWGLLDAG
801	801	KAMNGPASFP	FGDFTADTKG	TSDIAYSFRN	DISGTGGLIK	KGGSQLQLHG
851	851	NNTYTGKTII	EGGSLVLYGN	NKSDMRVETK	GALIYNGAAS	GGSLNSDGIV
901	901	YLADTDQSGA	NETVHIKGL	QLDGKGTLYT	RLGKLLKVDG	TAIIGGKLYM
951	951	SARGKGAGYL	NSTGRRVPFL	SAAKIGDYS	FFTNIETDGG	LLASLDSVEK
1001	1001	TAGSEGDTLS	YYVRRGNAAR	TASAAHSAP	AGLKHAVEQG	GSNLENLMVE
1051	1051	LDASESSATP	ETVETAAADR	TDMPGIRPYG	ATFRAAAAVQ	HANAADGVRI
1101	1101	FNSLAATVYA	DSTAHHADMQ	GRRLKAVSDG	LDHNGTGLRV	IAQTQQDGGT

1151 WEQGGVEGKMRGSTQTVGIA AKTGENTTAA ATLGMGRSTW SENSANAKTD
1201 SISLFAGIRH DAGDIGYLKG LFSYGRYKNS ISRSTGADEH AEGSVNGTLM
1251 QLGALGGVNV PFAATGDLTV EGGLRYDLLK QDAFAEKGSALGWGNSNLTE
1301 GTLVGLAGLK LSQPLSDKAV LFATAGVERD LNGRDYTVTG GFTGATAATG
1351 KTGARNMPHT RLVAGLGADV EFGNGWNGLA RYSYAGSKQY GNHSGRVGVG
1401 YRF*

It will be understood that the invention has been described by way of example only and modifications may be made whilst remaining within the scope and spirit of the invention. For instance, the use of proteins from other strains is envisaged [e.g. see WO00/66741 for polymorphic sequences for ORF4, ORF40, ORF46, 225, 235, 287, 519, 726, 919 and 953].

EXPERIMENTAL DETAILS

FPLC protein purification

15 The following table summarises the FPLC protein purification that was used:

Protein	PI	Column	Buffer	pH	Protocol
121.1 ^{untagged}	6.23	Mono Q	Tris	8.0	A
128.1 ^{untagged}	5.04	Mono Q	Bis-Tris propane	6.5	A
406.1L	7.75	Mono Q	Diethanolamine	9.0	B
576.1L	5.63	Mono Q	Tris	7.5	B
593 ^{untagged}	8.79	Mono S	Hepes	7.4	A
726 ^{untagged}	4.95	Hi-trap S	Bis-Tris	6.0	A
919 ^{untagged}	10.5(-leader)	Mono S	Bicine	8.5	C
919Lorf4	10.4(-leader)	Mono S	Tris	8.0	B
920L	6.92(-leader)	Mono Q	Diethanolamine	8.5	A
953L	7.56(-leader)	Mono S	MES	6.6	D
982 ^{untagged}	4.73	Mono Q	Bis-Tris propane	6.5	A
919-287	6.58	Hi-trap Q	Tris	8.0	A
953-287	4.92	Mono Q	Bis-Tris propane	6.2	A

Buffer solutions included 20-120 mM NaCl, 5.0 mg/ml CHAPS and 10% v/v glycerol. The dialysate was centrifuged at 13000g for 20 min and applied to either a mono Q or mono S FPLC ion-exchange resin. Buffer and ion exchange resins were chosen according to the pI of the protein of interest and the recommendations of the FPLC protocol manual [Pharmacia:

Publication]. Proteins were eluted using a step-wise NaCl gradient. Purification was analysed by SDS-PAGE and protein concentration determined by the Bradford method.

The letter in the 'protocol' column refers to the following:

FPLC-A: Clones 121.1, 128.1, 593, 726, 982, periplasmic protein 920L and hybrid proteins 5 919-287, 953-287 were purified from the soluble fraction of *E.coli* obtained after disruption of the cells. Single colonies harbouring the plasmid of interest were grown overnight at 37°C in 20 ml of LB/Amp (100 µg/ml) liquid culture. Bacteria were diluted 1:30 in 1.0 L of fresh medium and grown at either 30°C or 37°C until the OD₅₅₀ reached 0.6-08. Expression of recombinant protein was induced with IPTG at a final concentration of 1.0 mM. After 10 incubation for 3 hours, bacteria were harvested by centrifugation at 8000g for 15 minutes at 4°C. When necessary cells were stored at -20°C. All subsequent procedures were performed on ice or at 4°C. For cytosolic proteins (121.1, 128.1, 593, 726 and 982) and periplasmic protein 920L, bacteria were resuspended in 25 ml of PBS containing complete protease inhibitor (Boehringer-Mannheim). Cells were lysed by sonication using a Branson 15 Sonifier 450. Disrupted cells were centrifuged at 8000g for 30 min to sediment unbroken cells and inclusion bodies and the supernatant taken to 35% v/v saturation by the addition of 3.9 M (NH₄)₂SO₄. The precipitate was sedimented at 8000g for 30 minutes. The supernatant was taken to 70% v/v saturation by the addition of 3.9 M (NH₄)₂SO₄ and the precipitate collected as above. Pellets containing the protein of interest were identified by SDS-PAGE 20 and dialysed against the appropriate ion-exchange buffer (see below) for 6 hours or overnight. The periplasmic fraction from *E.coli* expressing 953L was prepared according to the protocol of Evans *et. al.* [Infect. Immun. (1974) 10:1010-1017] and dialysed against the appropriate ion-exchange buffer. Buffer and ion exchange resin were chosen according to the pI of the protein of interest and the recommendations of the FPLC protocol manual 25 (Pharmacia). Buffer solutions included 20 mM NaCl, and 10% (v/v) glycerol. The dialysate was centrifuged at 13000g for 20 min and applied to either a mono Q or mono S FPLC ion-exchange resin. Buffer and ion exchange resin were chosen according to the pI of the protein of interest and the recommendations of the FPLC protocol manual (Pharmacia). Proteins were eluted from the ion-exchange resin using either step-wise or continuous NaCl 30 gradients. Purification was analysed by SDS-PAGE and protein concentration determined by Bradford method. Cleavage of the leader peptide of periplasmic proteins was demonstrated by sequencing the NH₂-terminus (see below).

FPLC-B: These proteins were purified from the membrane fraction of *E.coli*. Single colonies harbouring the plasmid of interest were grown overnight at 37°C in 20 ml of LB/Amp (100 µg/ml) liquid culture. Bacteria were diluted 1:30 in 1.0 L of fresh medium. Clones 406.1L and 919LOrf4 were grown at 30°C and Orf25L and 576.1L at 37°C until the 5 OD₅₅₀ reached 0.6-0.8. In the case of 919LOrf4, growth at 30°C was essential since expression of recombinant protein at 37°C resulted in lysis of the cells. Expression of recombinant protein was induced with IPTG at a final concentration of 1.0 mM. After 10 incubation for 3 hours, bacteria were harvested by centrifugation at 8000g for 15 minutes at 4°C. When necessary cells were stored at -20 °C. All subsequent procedures were performed 15 at 4°C. Bacteria were resuspended in 25 ml of PBS containing complete protease inhibitor (Boehringer-Mannheim) and lysed by osmotic shock with 2-3 passages through a French Press. Unbroken cells were removed by centrifugation at 5000g for 15 min and membranes precipitated by centrifugation at 100000g (Beckman Ti50, 38000rpm) for 45 minutes. A Dounce homogenizer was used to re-suspend the membrane pellet in 7.5 ml of 20 mM Tris-HCl (pH 8.0), 1.0 M NaCl and complete protease inhibitor. The suspension was mixed for 2- 20 4 hours, centrifuged at 100000g for 45 min and the pellet resuspended in 7.5 ml of 20mM Tris-HCl (pH 8.0), 1.0M NaCl, 5.0mg/ml CHAPS, 10% (v/v) glycerol and complete protease inhibitor. The solution was mixed overnight, centrifuged at 100000g for 45 minutes and the supernatant dialysed for 6 hours against an appropriately selected buffer. In the case of 25 Orf25.L, the pellet obtained after CHAPS extraction was found to contain the recombinant protein. This fraction, without further purification, was used to immunise mice.

FPLC-C: Identical to FPLC-A, but purification was from the soluble fraction obtained after permeabilising *E.coli* with polymyxin B, rather than after cell disruption.

FPLC-D: A single colony harbouring the plasmid of interest was grown overnight at 37°C 25 in 20 ml of LB/Amp (100 µg/ml) liquid culture. Bacteria were diluted 1:30 in 1.0 L of fresh medium and grown at 30°C until the OD₅₅₀ reached 0.6-0.8. Expression of recombinant protein was induced with IPTG at a final concentration of 1.0mM. After incubation for 3 hours, bacteria were harvested by centrifugation at 8000g for 15. minutes at 4°C. When necessary cells were stored at -20 °C. All subsequent procedures were performed on ice or at 30 4°C. Cells were resuspended in 20mM Bicine (pH 8.5), 20mM NaCl, 10% (v/v) glycerol, complete protease inhibitor (Boehringer-Mannheim) and disrupted using a Branson Sonifier 450. The sonicate was centrifuged at 8000g for 30 min to sediment unbroken cells and

inclusion bodies. The recombinant protein was precipitated from solution between 35% v/v and 70% v/v saturation by the addition of 3.9M $(\text{NH}_4)_2\text{SO}_4$. The precipitate was sedimented at 8000g for 30 minutes, resuspended in 20 mM Bicine (pH 8.5), 20 mM NaCl, 10% (v/v) glycerol and dialysed against this buffer for 6 hours or overnight. The dialysate was 5 centrifuged at 13000g for 20 min and applied to the FPLC resin. The protein was eluted from the column using a step-wise NaCl gradients. Purification was analysed by SDS-PAGE and protein concentration determined by Bradford method.

Cloning strategy and oligonucleotide design

Genes coding for antigens of interest were amplified by PCR, using oligonucleotides 10 designed on the basis of the genomic sequence of *N. meningitidis* B MC58. Genomic DNA from strain 2996 was always used as a template in PCR reactions, unless otherwise specified, and the amplified fragments were cloned in the expression vector pET21b+ (Novagen) to express the protein as C-terminal His-tagged product, or in pET-24b+(Novagen) to express the protein in 'untagged' form (e.g. ΔG 287K).

15 Where a protein was expressed without a fusion partner and with its own leader peptide (if present), amplification of the open reading frame (ATG to STOP codons) was performed.

Where a protein was expressed in 'untagged' form, the leader peptide was omitted by designing the 5'-end amplification primer downstream from the predicted leader sequence.

20 The melting temperature of the primers used in PCR depended on the number and type of hybridising nucleotides in the whole primer, and was determined using the formulae:

$$T_{m1} = 4 (\text{G}+\text{C}) + 2 (\text{A}+\text{T}) \quad (\text{tail excluded})$$

$$T_{m2} = 64.9 + 0.41 (\% \text{ GC}) - 600/\text{N} \quad (\text{whole primer})$$

The melting temperatures of the selected oligonucleotides were usually 65-70°C for the whole oligo and 50-60°C for the hybridising region alone.

25 Oligonucleotides were synthesised using a Perkin Elmer 394 DNA/RNA Synthesizer, eluted from the columns in 2.0ml NH₄OH, and deprotected by 5 hours incubation at 56°C. The oligos were precipitated by addition of 0.3M Na-Acetate and 2 volumes ethanol. The samples were centrifuged and the pellets resuspended in water.

		Sequences	Restriction site
Orf1L	Fwd	CGCGGATCCGCTAGC-AAAACAACCGACAAACGG	NheI
	Rev	CCCG <u>CTCGAG</u> -TTACCAGCGGTAGCCTA	XhoI
Orf1	Fwd	CTAG <u>CTAGC</u> -GGACACACTTATTCGGCATC	NheI
	Rev	CCCG <u>CTCGAG</u> -TTACCAGCGGTAGCCTAATTG	XhoI
Orf1LOmpA	Fwd		NdeI-(NheI)
	Rev	CCCG <u>CTCGAG</u> -	XhoI
Orf4L	Fwd	CGCGGAT <u>CCCATATG</u> -AAAACCTTCTCAAAACC	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTATTGCGCTGCCTTC	XhoI
Orf7-1L	Fwd	GCGG <u>CATTAAT</u> -ATGTTGAGAAAATTGTTGAAATGG	AseI
	Rev	GCGG <u>CTCGAG</u> -TTATTTTCAAAATATATITGC	XhoI
Orf9-1L	Fwd	GCGG <u>CATATG</u> -TTACCTAACCGTTCAAAATGT	NdeI
	Rev	GCGG <u>CTCGAG</u> -TTATT <u>CCGAGGTTTCGGG</u>	XhoI
Orf23L	Fwd	CGCGGAT <u>CCCATATG</u> -ACACGCTTCAAATATT	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTATTAAACCGATAGGTAAA	XhoI
Orf25-1 His	Fwd	CGCGGAT <u>CCCATATG</u> -GGCAGGGAAAGAACCGC	NdeI
	Rev	GCCC <u>AAGCTT</u> -ATCGATGGAATAGCCCG	HindIII
Orf29-1 b-His (MC58)	Fwd	CGCGGAT <u>CCGCTAGC</u> -AACGGTTGGATGCCCG	NheI
	Rev	CCCG <u>CTCGAG</u> -TTTGTCTAACGTTCTGATAT CCCG <u>CTCGAG</u> -ATTCCACCTGCCATC	XhoI
Orf29-1 b-L (MC58)	Fwd	CGCGGAT <u>CCGCTAGC</u> -ATGAATTGCTATTCAAAAT	NheI
	Rev	CCCG <u>CTCGAG</u> -TTAATT <u>CCCACCTGCCATC</u>	XhoI
Orf29-1 c-His (MC58)	Fwd	CGCGGAT <u>CCGCTAGC</u> -ATGAATTGCTATTCAAAAT	NheI
	Rev	CCCG <u>CTCGAG</u> -TTGGACGATGCCCGCGA	XhoI
Orf29-1 c-L (MC58)	Fwd	CGCGGAT <u>CCGCTAGC</u> -ATGAATTGCTATTCAAAAT	NheI
	Rev	CCCG <u>CTCGAG</u> -TTATTGGACGATGCCCGC	XhoI
Orf25L	Fwd	CGCGGAT <u>CCCATATG</u> -TATCGCAA <u>CTGATTGC</u>	NdeI
	Rev	CCCG <u>CTCGAG</u> -CTAATCGATGGAATAGCC	XhoI
Orf37L	Fwd	CGCGGAT <u>CCCATATG</u> -AAACAGACAGTCAAATG	NdeI
	Rev	CCCG <u>CTCGAG</u> -TCAATAACCGCCTTCAG	XhoI
Orf38L	Fwd	CGCGGAT <u>CCCATATG</u> -TTACGTTGACTGCTTAGCCGTATGCACC	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTATTGCGCGTAAAAGCGTCGGCAAC	XhoI
Orf40L	Fwd	CGCGGAT <u>CCCATATG</u> -AACAAAATATACCGCAT	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTACCACTGATAACCGAC	XhoI
Orf40.2-His	Fwd	CGCGGAT <u>CCCATATG</u> -ACCGATGACGACGATTAT	NdeI
	Rev	GCCC <u>AAGCTT</u> -CCACTGATAACCGACAGA	HindIII
Orf40.2L	Fwd	CGCGGAT <u>CCCATATG</u> -AACAAAATATACCGCAT	NdeI
	Rev	GCCC <u>AAGCTT</u> -TTACCACTGATAACCGAC	HindIII
Orf46-2L	Fwd	GGGAATT <u>CCCATATG</u> -GGCATTCCCGCAAAATATC	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTATT <u>ACTCCTATAACGAGGTCTCTAAC</u>	XhoI
Orf46-2	Fwd	GGGAATT <u>CCCATATG</u> -TCAGATTGGCAAACGATTCTT	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTATT <u>ACTCCTATAACGAGGTCTCTAAC</u>	XhoI
Orf46.1L	Fwd	GGGAATT <u>CCCATATG</u> -GGCATTCCCGCAAAATATC	NdeI

	Rev	CCCGCTCGAG-TTACGTATCATATTCACGTGC	XhoI
orf46. (His-GST)	Fwd	GGGAATTCCATATGCACGTGAAATATGATACGAAG	BamHI-NdeI
	Rev	CCCGCTCGAGTTACTCCTATAACGAGGTCTCTAAC	XhoI
orf46.1-His	Fwd	GGGAATTCCATATGTCAGATTGGCAAACGATTCTT	NdeI
	Rev	CCCGCTCGAGCGTATCATATTCACGTGC	XhoI
orf46.2-His	Fwd	GGGAATTCCATATGTCAGATTGGCAAACGATTCTT	NdeI
	Rev	CCCGCTCGAGTTACTCCTATAACGAGGTCTCTAAC	XhoI
Orf65-1-(His/GST) (MC58)	Fwd	CGCGGATCCCATATG-CAAAATGCGTCAAAATCCC	BamHI-NdeI
	Rev	CGCGGATCCCATATG-AACAAAATATACCGCAT CCCGCTCGAG -TTGCTTCGATAGAACCG	XhoI
Orf72-1L	Fwd	CGGGCCATATG-GTCATAAAATATACAAATTGAA	NdeI
	Rev	CGGGCCTCGAG-TTACGGCTGAGACCTTTGCAAATT	XhoI
Orf76-1L	Fwd	CGGGCCATATG-AAACAGAAAAAAACCGCTG	NdeI
	Rev	CGGGCCTCGAG-TTACGGTTGACACCGTTTC	XhoI
Orf83.1L	Fwd	CGCGGATCCCATATG-AAAACCCCTGCTCCTC	NdeI
	Rev	CCCGCTCGAG-TTATCCTCCTTGCGGC	XhoI
Orf85-2L	Fwd	CGGGCCATATG-GCAAAATGATGAAATGGG	NdeI
	Rev	CGGGCCTCGAG-TTATCGCGCGGGCGGGCC	XhoI
Orf91L (MC58)	Fwd	CGGGCCATATGAAAAAAATCCTCCCTCATCA	NdeI
	Rev	CGGGCCTCGAGTTATTCGCCGCCGTTTGGC	XhoI
Orf91-His(MC58)	Fwd	CGGGCCATATGGCCCTGCGGACGCGGTAAG	NdeI
	Rev	CGGGCCTCGAGTTGCCGCCGTTTGGCTTTC	XhoI
Orf97-1L	Fwd	CGGGCCATATG-AAACACATACTCCCCCTGA	NdeI
	Rev	CGGGCCTCGAG-TTATCGCCTACGGTTTTG	XhoI
Orf119L (MC58)	Fwd	CGGGCCATATGATTACATCGTACTGTTTC	NdeI
	Rev	CGGGCCTCGAGTTAGGAGAACAGGCGCAATGC	XhoI
Orf119-His(MC58)	Fwd	CGGGCCATATGTACAACATGTATCAGGAAAAC	NdeI
	Rev	CGGGCCTCGAGGGAGAACAGGCGCAATGCGG	XhoI
Orf137.1 (His-GST) (MC58)	Fwd	CGCGGATCCGCTAGCTGCCGCACGGCGGG	BamHI-NheI
	Rev	CCCGCTCGAGATAACGGTATGCCGCCAG	XhoI
Orf143-1L	Fwd	CGCGGATCCCATATG-GAATCAACACTTCAC	NdeI
	Rev	CCCGCTCGAG-TTACACGCGGGTTGCTGT	XhoI
008	Fwd	CGCGGATCCCATATG-AACAAACAGACATTG	NdeI
	Rev	CCCGCTCGAG-TTACCTGTCCGGTAAAG	XhoI
050-1(48)	Fwd	CGCGGATCCGCTAGC-ACCGTCATCAAACAGGAA	NheI
	Rev	CCCGCTCGAG-TCAAGATTGACGGGGA	XhoI
105	Fwd	CGCGGATCCCATATG-TCCGCAAACGAATACG	NdeI
	Rev	CCCGCTCGAG-TCAGTGTCTGCCAGTT	XhoI
111L	Fwd	CGCGGATCCCATATG-CCGTCTGAAACACG	NdeI
	Rev	CCCGCTCGAG-TTACGGAGCAGTTTTC	XhoI
117-1	Fwd	CGCGGATCCCATATG-ACGCCATCAGCC	NdeI
	Rev	CCCGCTCGAG-TTAAAGCCGGGTAACGC	XhoI
121-1	Fwd	CGGGCCATATG-GAAACACAGCTTACATCGG	NdeI
	Rev	CGGGCCTCGAG-TCAATAATAATCCCGCG	XhoI

122-1	Fwd	CGGGCCATATG-ATTAATCCGCAATATCC	NdeI
	Rev	CGGGCCTCGAG-TTAAATCTGGTAGATTGGATTGG	XhoI
128-1	Fwd	CGGGCCATATG-ACTGACAACGCACTGCTCC	NdeI
	Rev	CGGGCCTCGAG-TCAGACCGCGTTGTCGAAAC	XhoI
148	Fwd	CGCGGATCCCATATG-GCGTTAAAAACATCAA	NdeI
	Rev	CCCGCTCGAG-TCAGCCCTTCATACAGC	XhoI
149.1L (MC58)	Fwd	CGGGCATTAATGGCACAAACTACACTCAAACC	AseI
	Rev	CGGGCCTCGAGTTAAAACCTCACGTTCACGCCG	XhoI
149.1-His(MC58)	Fwd	CGGGCATTAATGCATGAAACTGAGCAATCGGTGG	AseI
	Rev	CGGGCCTCGAGAAACCTTCACGTTCACGCCGCCGTAAA	XhoI
205 (His-GST) (MC58)	Fwd	CGCGGATCCCATATGGCAAATCCGAAAATACG	BamHI-NdeI
	Rev	CCCGCTCGAGATAATGGCGGCCGG	XhoI
206L	Fwd	CGCGGATCCCATATG-TTCCCCCGACAA	NdeI
	Rev	CCCGCTCGAG-TCATTCTGTAAAAAAAGTATG	XhoI
214 (His-GST) (MC58)	Fwd	CGCGGATCCCATATGCTTCAAAGCGACAGCAG	BamHI-NdeI
	Rev	CCCGCTCGAGTTCGGATTTGCGTACTC	XhoI
216	Fwd	CGCGGATCCCATATG-GCAATGGCAGAAAACG	NdeI
	Rev	CCCGCTCGAG-CTATACAATCCGTGCCG	XhoI
225-1L	Fwd	CGCGGATCCCATATG-GATTCTTTCAAACC	NdeI
	Rev	CCCGCTCGAG-TCAGTTCAGAAAGCGGG	XhoI
235L	Fwd	CGCGGATCCCATATG-AAACCTTGATTAGG	NdeI
	Rev	CCCGCTCGAG-TTATTGGGCTGCTCTTC	XhoI
243	Fwd	CGCGGATCCCATATG-GTAATCGTCTGGTTG	NdeI
	Rev	CCCGCTCGAG-CTACGACTTGGTACCG	XhoI
247-1L	Fwd	CGGGCCATATG-AGACGTAATGCTAAAGCTAC	NdeI
	Rev	CGGGCCTCGAG-TCAAAGTGTCTGTTGCGC	XhoI
264-His	Fwd	CGGGCCATATG-TTGACTTAACCGAAAAAA	NdeI
	Rev	CGCGCTCGAG-GCCGGCGGTCAATACCGCCCCGAA	XhoI
270 (His-GST) (MC58)	Fwd	CGCGGATCCCATATGGCGCAATCGGATTGAC	BamHI-NdeI
	Rev	CCCGCTCGAGTTCGGCGGTAAATGCCG	XhoI
274L	Fwd	CGGGCCATATG-GCGGGGCCGATTTGT	NdeI
	Rev	CGGGCCTCGAG-TTATTGCTTCAGTATTATTG	XhoI
283L	Fwd	CGGGCCATATG-AACTTGCTTATCCGTCA	NdeI
	Rev	CGGGCCTCGAG-TTAACGGCAGTATTGTTAC	XhoI
285-His	Fwd	CGCGGATCCCATATGGTTGCGCTTCGGGC	BamHI
	Rev	GCCCAAGCTTTTCCCGTTGCCGTTCCG	HindIII
286-His (MC58)	Fwd	CGCGGATCCCATATG-GCCGACCTTCCGAAAAA	NdeI
	Rev	CCCGCTCGAG-GAAGCGCGTCCCAAGC	XhoI
286L (MC58)	Fwd	CGCGGATCCCATATG-CACGACACCCGTAC	NdeI
	Rev	CCCGCTCGAG-TTACAAGCGCGTCCCAA	XhoI
287L	Fwd	CTAGCTAGC-TTAAACGCGAGCGTAATCGCAATGG	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTGCC	XhoI

287	Fwd	CTAGCTAGC-GGGGGCGGCGGGTGGCG	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTGCC	XhoI
287LOrf4	Fwd	CTAGCTAGCGCTATCCTCGCCGCC-TGCGGGGGCGGCGGT	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTGCC	XhoI
287-fu	Fwd	CGGGGATCC-GGGGGCGGCGGGTGGCG	BamHI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTGCC	XhoI
287-His	Fwd	CTAGCTAGC-GGGGGCGGCGGGTGGCG	NheI
	Rev	CCCGCTCGAG-ATCCTGCTCTTTTGCC *	XhoI
287-His(2996)	Fwd	CTAGCTAGC-TGCGGGGGCGGCGGTGGCG	NheI
	Rev	CCCGCTCGAG-ATCCTGCTCTTTTGCC	XhoI
Δ1 287-His	Fwd	CGCGGATCCGCTAGC-CCCGATGTTAAATCGGC §	NheI
Δ2 287-His	Fwd	CGCGGATCCGCTAGC-CAAGATATGGCGGCAGT §	NheI
Δ3 287-His	Fwd	CGCGGATCCGCTAGC-GCCGAATCCGCAAATCA §	NheI
Δ4 287-His	Fwd	CGCGCTAGC-GGAAGGGTTGATTGGCTAATGG §	NheI
Δ4 287MC58-His	Fwd	CGCGCTAGC-GGAAGGGTTGATTGGCTAATGG §	NheI
287a-His	Fwd	CGCCATATG-TTAAACGCAACGTAATCGC	NdeI
	Rev	CCCGCTCGAG-AAAATTGCTACCGCCATTGCAAGG	XhoI
287b-His	Fwd	CGCCATATG-GGAAGGGTTGATTGGCTAATGG	NdeI
287b-2996-His	Rev	CCCGCTCGAG-CTTGTCTTATAATGATGACATATTG	XhoI
287b-MC58-His	Rev	CCCGCTCGAG-TTTATAAAAGATAATATATTGATTGATTCC	XhoI
287c-2996-His	Fwd	CGCGCTAGC-ATGCCGCTGATTCCGCTAACTC §	NheI
‘287^{untagged}(2996)	Fwd	CTAGCTAGC-GGGGGCGGCGGGTGGCG	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTGCC	XhoI
ΔG287-His *	Fwd	CGCGGATCCGCTAGC-CCCGATGTTAAATCGGC	NheI
	Rev	CCCGCTCGAG-ATCCTGCTCTTTTGCC	XhoI
ΔG287K(2996)	Fwd	CGCGGATCCGCTAGC-CCCGATGTTAAATCGGC	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTGCC	XhoI
ΔG 287-L	Fwd	CGCGGATCCGCTAGC-TTGAACGCAAGTGTGATTGCAATGGCTGTATTGGCTCAGCCTGT TCGCCCGATGTTAAATCGGC	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTGCC	XhoI
ΔG 287-Orf4L	Fwd	CGCGGATCCGCTAGC-AAAACCTTCTCAAAACCCCTTCCGCCGCCACTCGCG CTCATCCTGCCGCCTGC TCGCCCGATGTTAAATCGG	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTGCC	XhoI
292L	Fwd	CGCGGATCCCATATG-AAAACCAAGTAAATCAA	NdeI
	Rev	CCCGCTCGAG-TTATTGATTGGCTGGATGA	XhoI
308-1	Fwd	CGCGGATCCCATATG-TTAAATCGGTATTTATC	NdeI
	Rev	CCCGCTCGAG-TTAAATCGCCATTCCCTG	XhoI
401L	Fwd	GCGGCCATATG-AAATTACAACAATTGGCTG	NdeI
	Rev	GCGGCCCTCGAG-TTACCTACGTTTCAAAG	XhoI
406L	Fwd	CGCGGATCCCATATG-CAAGCACGGCTGCT	NdeI
	Rev	CCCGCTCGAG-TCAAGGTTGTCCTTGTCTA	XhoI
502-1L	Fwd	CGCGGATCCCATATG-ATGAAACCGCACAAC	NdeI
	Rev	CCCGCTCGAG-TCAGTTGCTAACACAGTC	XhoI

502-A (His-GST)	Fwd	CGCGGATCCCATATGGTAGACGCGCTTAAGCA	BamHI-NdeI
	Rev	CCCGCTCGAGAGCTGCATGGCGCG	XhoI
503-1L	Fwd	CGCGGATCCCATATG-GCACGGTCGTATAC	NdeI
	Rev	CCCGCTCGAG-CTACCGCGCATTCTG	XhoI
519-1L	Fwd	GC GG CATATG-GAATTTTCA TT CTGTT	NdeI
	Rev	GC GG C CT CGAG-TTATTTGGCGGTTTGCTGC	XhoI
525-1L	Fwd	GC GG CATATG-AAGTATGTCCGGTTATTTTC	NdeI
	Rev	GC GG C CT CGAG-TTATCGGTTGTGCAACGG	XhoI
529-(His/GST) (MC58)	Fwd	CGCGGATCCGCTAGC-TCCGGCAGCAAAACCGA	Bam HI-NheI
	Rev	GCCCAAGCTT-ACGCAGTTCCGAATGGAG	HindIII
552L	Fwd	GCCGCCATATGTTGAATATTAAACTGAAAACCTG	NdeI
	Rev	GCCGCCTCGAGTTATTCGATGCCTTTCCC	XhoI
556L	Fwd	GCCGCCATATGGACAATAAGACCAAACGT	NdeI
	Rev	GCCGCCTCGAGTTAACGGTGCGGACGTTTC	XhoI
557L	Fwd	CGCGGATCCCATATG-AACAAACTGTTCTTAC	NdeI
	Rev	CCCGCTCGAG-TCATTCCGCCTTCAGAAA	XhoI
564ab-(His/GST) (MC58)	Fwd	CGCGGATCCCATATG-CAAGGTATCGTTGCCGACAAATCCGCACCT	BamHI-NdeI
	Rev	CCCGCTCGAG-AGCTAATTGTGCTTGGTTGCAGATAGGAGTT	XhoI
564abL (MC58)	Fwd	CGCGGATCCCATATG-AACCGCACCTGTACAAAGTTGATTAAACAAACATC	NdeI
	Rev	CCCGCTCGAG-TTAAGCTAATTGTGCTTGGTTGCAGATAGGAGTT	XhoI
564b-(His/GST)(MC58)	Fwd	CGCGGATCCCATATG-ACGGGAGAAAATCATGCGGTTCACTTCATG	BamHI-NdeI
	Rev	CCCGCTCGAG-AGCTAATTGTGCTTGGTTGCAGATAGGAGTT	XhoI
564c-(His/GST)(MC58)	Fwd	CGCGGATCCCATATG-GTTTCAGACGGCCTATACAACCAACATGGTGAAATT	BamHI-NdeI
	Rev	CCCGCTCGAG-GCGGTAACTGCCGCTTGCACTGAATCCGTAA	XhoI
564bc-(His/GST)(MC58)	Fwd	CGCGGATCCCATATG-ACGGGAGAAAATCATGCGGTTCACTTCATG	BamHI-NdeI
	Rev	CCCGCTCGAG-GCGGTAACTGCCGCTTGCACTGAATCCGTAA	XhoI
564d-(His/GST)(MC58)	Fwd	CGCGGATCCCATATG-CAAAGCAAAGTCAAAGCAGACCATGCCTCCGTAA	BamHI-NdeI
	Rev	CCCGCTCGAG-TCTTTCCCTTCAATTATAACTTTAGTAGGTTCAATTGGTGTCCCC	XhoI
564cd-(His/GST)(MC58)	Fwd	CGCGGATCCCATATG-GTTTCAGACGGCCTATACAACCAACATGGTGAAATT	BamHI-NdeI
	Rev	CCCGCTCGAG-TCTTTCCCTTCAATTATAACTTTAGTAGGTTCAATTGGTGTCCCC	XhoI
570L	Fwd	GC GG CATATG-ACCCGTTGACCCGCG	NdeI
	Rev	GC GG C CT CGAG-TCAGCGGGCGTTCA TT CTT	XhoI
576-1L	Fwd	CGCGGATCCCATATG-AACACCATTTCAAAATC	NdeI
	Rev	CCCGCTCGAG-TTAATTACTTTGTGTCG	XhoI

580L	Fwd	<u>GCGGCCATATG-GATTGCCCCAAGGTGG</u>	NdeI
	Rev	<u>GCGGCCTCGAG-CTACACTTCCCCGAAGTGG</u>	XhoI
583L	Fwd	<u>CGCGGATCCCATATG-ATAGTTGACCAAAGCC</u>	NdeI
	Rev	<u>CCCGCTCGAG-TTATTTTCCGATTTTCGG</u>	XhoI
593	Fwd	<u>GCGGCCATATG-CTTGAACGTAAACGGACT</u>	NdeI
	Rev	<u>GCGGCCTCGAG-TCAGCGGAAGCGGACGATT</u>	XhoI
650 (His-GST) (MC58)	Fwd	<u>CGCGGATCCCATATGTCACAAACTCAAAACCATCG</u>	BamHI-NdeI
	Rev	<u>CCCGCTCGAGGCTTCCAATCAGTTGACC</u>	XhoI
652	Fwd	<u>GCGGCCATATG-AGCGCAATCGTTGATATTTTC</u>	NdeI
	Rev	<u>GCGGCCTCGAG-TTATTTGCCAGTTGGTAGAATG</u>	XhoI
664L	Fwd	<u>GCGGCCATATG-GTGATACATCCGCACTACTTC</u>	NdeI
	Rev	<u>GCGGCCTCGAG-TCAAAATCGAGTTTACACCA</u>	XhoI
726	Fwd	<u>GCGGCCATATG-ACCATCTATTCACAAACGG</u>	NdeI
	Rev	<u>GCGGCCTCGAG-TCAGCCGATGTTAGCGTCCATT</u>	XhoI
741-His(MC58)	Fwd	<u>CGCGGATCCCATATG-AGCAGCGGAGGGGGTG</u>	NdeI
	Rev	<u>CCCGCTCGAG-TTGCTTGGCGGCAAGGC</u>	XhoI
ΔG741-His(MC58)	Fwd	<u>CGCGGATCCCATATG-GTCGCCGCCACATCG</u>	NdeI
	Rev	<u>CCCGCTCGAG-TTGCTTGGCGGCAAGGC</u>	XhoI
686-2-(His/GST) (MC58)	Fwd	<u>CGCGGATCCCATATG-GGCGGTTCGGAAGGCG</u>	BamHI-NdeI
	Rev	<u>CCCGCTCGAG-TTGAACACTGATGTCTTTCCGA</u>	XhoI
719-(His/GST) (MC58)	Fwd	<u>CGCGGATCCGCTAGC-AAACTGTCGTTGGTGTAAAC</u>	BamHI-NheI
	Rev	<u>CCCGCTCGAG-TTGACCCGCTCCACGG</u>	XhoI
730-His (MC58)	Fwd	<u>GCGGCCATATGGCGGACTTGGCGCAAGACCC</u>	NdeI
	Rev	<u>GCGGCCTCGAGATCTCCTAAACCTGTTAACATGCCG</u>	XhoI
730A-His (MC58)	Fwd	<u>GCGGCCATATGGCGGACTTGGCGCAAGACCC</u>	NdeI
	Rev	<u>GCGGCCTCGAGCTCATGCTGTTGCCAGC</u>	XhoI
730B-His (MC58)	Fwd	<u>GCGGCCATATGGCGGACTTGGCGCAAGACCC</u>	NdeI
	Rev	<u>GCGGCCTCGAGAAAATCCCCGCTAACCGCAG</u>	XhoI
741-His (MC58)	Fwd	<u>CGCGGATCCCATATG-AGCAGCGGAGGGGGTG</u>	NdeI
	Rev	<u>CCCGCTCGAG-TTGCTTGGCGGCAAGGC</u>	XhoI
ΔG741-His (MC58)	Fwd	<u>CGCGGATCCCATATG-GTCGCCGCCACATCG</u>	NdeI
	Rev	<u>CCCGCTCGAG-TTGCTTGGCGGCAAGGC</u>	XhoI
743 (His-GST)	Fwd	<u>CGCGGATCCCATATGGACGGTGTGCGCTGTT</u>	BamHI-NdeI
	Rev	<u>CCCGCTCGAGCTTACGGATCAAATTGACG</u>	XhoI
757 (His-GST) (MC58)	Fwd	<u>CGCGGATCCCATATGGCAGCCAATCTGAAGAA</u>	BamHI-NdeI
	Rev	<u>CCCGCTCGAGCTCAGCTTGTGCCGTCAA</u>	XhoI
759-His/GST (MC58)	Fwd	<u>CGCGGATCCGCTAGC-TACTCATCCATTGTCCGC</u>	BamHI-NheI
	Rev	<u>CCCGCTCGAG-CCAGTTGTAGCCTATTTG</u>	XhoI
759L (MC58)	Fwd	<u>CGCGGATCCGCTAGC-ATGCGCTTCACACACAC</u>	NheI
	Rev	<u>CCCGCTCGAG-TTACCAAGTTGTAGCCTATTT</u>	XhoI
760-His	Fwd	<u>GCGGCCATATGGCACAAACGGAAGGTTGGAA</u>	NdeI
	Rev	<u>GCGGCCTCGAGAAAATGTAACGCAGGTTGCCGTC</u>	XhoI
769-His (MC58)	Fwd	<u>GCGGCCATATGGAAGAACACCGCGCGAACCG</u>	NdeI

	Rev	GGGGCCTCGAGGAACGTTTATTAAACTCGAC	XhoI
907L	Fwd	GGGCC <u>CATATG</u> -AGAAAACCGACCGATACCTA	NdeI
	Rev	GGGG <u>CCTCGAG</u> -TCAACGCCACTGCCAGCGGTG	XhoI
911L	Fwd	CGGGAT <u>CCCATATG</u> -AAGAAGAACATATTGAATTTGGTCCGACTG	NdeI
	Rev	CCC <u>GCTCGAG</u> -TTATCGCGGCTTTCCGCATTGCCG	XhoI
911LOmpA	Fwd	GGGAATT <u>CCATATG</u> AAAAAGACAGCTATCGCGATTGCA GTGGCACTGGCTGGTTCGCTACCGTAGCGCAGGCC <u>GC</u> <u>TAGC</u> -GCTTCCGCGTGGCCGGTGC	NdeI-(NheI)
	Rev	CCC <u>GCTCGAG</u> -TTATCGCGGCTTTCCGCATTGCCG	XhoI
911LPelB	Fwd	CAT <u>GCATGG</u> -CTTCCGCGTGGCCGGCGGTGC	NcoI
	Rev	CCC <u>GCTCGAG</u> -TTATCGCGGCTTTCCGCATTGCCG	XhoI
913-His/GST (MC58)	Fwd	CGCGGAT <u>CCCATATG</u> -TTGCCGAAACCCGCC	BamHI-NdeI
	Rev	CCC <u>GCTCGAG</u> -AGGTTGTGTCAGGTG	XhoI
913L (MC58)	Fwd	CGGGAT <u>CCCATATG</u> -AAAAAAACCGCCTATG	NdeI
	Rev	CCC <u>GCTCGAG</u> -TTAAGGTTGTGTCAGG	XhoI
919L	Fwd	CGGGAT <u>CCCATATG</u> -AAAAAAACCTATTCCGC	NdeI
	Rev	CCC <u>GCTCGAG</u> -TTACGGCGGTATTGG	XhoI
919	Fwd	CGGGAT <u>CCCATATG</u> -CAAAGCAAGAGCATCCAA	NdeI
	Rev	CCC <u>GCTCGAG</u> -TTACGGCGGTATTGG	XhoI
919L Orf4	Fwd	GGGAATT <u>CCATATG</u> AAAACCTCTTCAAAACCCCTTCCG CCGCG <u>CGCTAGCGCTCATCCTCGCCGCC</u> TGCAAAGCAAGAGCATC	NdeI-(NheI)
	Rev	CCC <u>GCTCGAG</u> -TTACGGCGGTATTGGCTTCATACCG	XhoI
(919)-287fusion	Fwd	CGCGGAT <u>CCGTCGAC</u> -TGTGGGGCGGGTGGC	SalI
	Rev	CCC <u>GCTCGAG</u> -TCAATCCTGCTTTTGCC	XhoI
920-1L	Fwd	GG <u>GCATATG</u> -AAGAAAACATTGACACTGC	NdeI
	Rev	GG <u>GCCTCGAG</u> -TTAATGGTGCAGGAT	XhoI
925-His/GST (MC58) ^{GATE}	Fwd	ggggacaagtgtacaaaaaagcaggctTGC <u>GGCAAGGATGCCGG</u>	<i>attB1</i>
	Rev	ggggaccacttgtacaagaaagctgggt <u>CTAAAGCAACAATGCCGG</u>	<i>attB2</i>
926L	Fwd	CGGGAT <u>CCCATATG</u> -AAACACACCGTATCC	NdeI
	Rev	CCC <u>GCTCGAG</u> -TTATCTGTGCGCGCC	XhoI
927-2-(His/GST) (MC58)	Fwd	CG <u>GGATCCCATATG</u> -AGCCCCGCGCCGATT	BamHI-NdeI
	Rev	CCC <u>GCTCGAG</u> -TTTTGTGCGGT <u>CAGGCG</u>	XhoI
932-His/GST (MC58) ^{GATE}	Fwd	ggggacaagtgtacaaaaaagcaggct <u>TGTCGTTGGGGATTAA</u> ACCAAA <u>CCAAATC</u>	<i>attB1</i>
	Rev	CG <u>GGATCCCATATGGCGGATGCGCCCGCG</u>	BamHI-NdeI
935 (His-GST) (MC58)	For	CCC <u>GCTCGAGAAACGCCAATCCGCC</u>	XhoI
	Rev	ggggaccacttgtacaagaaagctgggt <u>TCATTTGTTTCCCTTCT</u> CGAGGCCATT	<i>attB2</i>
936-1L	Fwd	CGGG <u>ATCCCATATG</u> -AAACCCAAACCGCAC	NdeI
	Rev	CCC <u>GCTCGAG</u> -TCAGCGTTGGACGTAGT	XhoI
953L	Fwd	GGGAATT <u>CCATATG</u> -AAAAAAATCATCTCGCCG	NdeI
	Rev	CCC <u>GCTCGAG</u> -TTATTGTTGGCTGCCTCGAT	XhoI
953-fu	Fwd	GGGAATT <u>CCATATG</u> -GCCACCTACAAAGTGGACG	NdeI
	Rev	CGGG <u>GATCC</u> -TTGTTGGCTGCCTCGATTG	BamHI

954 (His-GST) (MC58)	Fwd	CGCGGATCCCATATGCAAGAACAAATCGCAGAAAG	BamHI-NdeI
	Rev	CCCGCTCGAGTTTTTCGGCAAATTGGCTT	XhoI
958-His/GST (MC58) ^{GATE}	Fwd	ggggacaagttgtacaaaaaaaggcaggctGCCGATGCCGTTGCGG	attB1
	Rev	ggggaccacttgcataagaaagctgggtTCAGGGTCGTTGCG	attB2
961L	Fwd	CGCGGATCCCATATG-AAACACTTCCATCC	NdeI
	Rev	CCCGCTCGAG-TTACCACTCGTAATTGAC	XhoI
961	Fwd	CGCGGATCCCATATG-GCCACAAGCGACGAC	NdeI
	Rev	CCCGCTCGAG-TTACCACTCGTAATTGAC	XhoI
961 c (His/GST)	Fwd	CGCGGATCCCATATG-GCCACAAACGACG	BamHI-NdeI
	Rev	CCCGCTCGAG-ACCCACGTTGTAAGGTTG	XhoI
961 c-(His/GST) (MC58)	Fwd	CGCGGATCCCATATG-GCCACAAGCGACGACGA	BamHI-NdeI
	Rev	CCCGCTCGAG-ACCCACGTTGTAAGGTTG	XhoI
961 c-L	Fwd	CGCGGATCCCATATG-ATGAAACACTTCCATCC	NdeI
	Rev	CCCGCTCGAG-TTAACCCACGTTGTAAGGT	XhoI
961 c-L (MC58)	Fwd	CGCGGATCCCATATG-ATGAAACACTTCCATCC	NdeI
	Rev	CCCGCTCGAG-TTAACCCACGTTGTAAGGT	XhoI
961 d (His/GST)	Fwd	CGCGGATCCCATATG-GCCACAAACGACG	BamHI-NdeI
	Rev	CCCGCTCGAG-GTCTGACACTGTTTATCC	XhoI
961 Δ1-L	Fwd	CGCGGATCCCATATG-ATGAAACACTTCCATCC	NdeI
	Rev	CCCGCTCGAG-TTATGCTTGGCGGCAAAG	XhoI
fu 961-...	Fwd	CGCGGATCCCATATG- GCCACAAACGACGAC	NdeI
	Rev	CGCGGATCC-CCACTCGTAATTGACGCC	BamHI
fu 961-... (MC58)	Fwd	CGCGGATCCCATATG-GCCACAAGCGACGAC	NdeI
	Rev	CGCGGATCC-CCACTCGTAATTGACGCC	BamHI
fu 961 c -...	Fwd	CGCGGATCCCATATG-GCCACAAACGACGAC	NdeI
	Rev	CGCGGATCC -ACCCACGTTGTAAGGTTG	BamHI
fu 961 c-L-...	Fwd	CGCGGATCCCATATG- ATGAAACACTTCCATCC	NdeI
	Rev	CGCGGATCC -ACCCACGTTGTAAGGTTG	BamHI
fu (961)- 741(MC58)-His	Fwd	CGCGGATCC -GGAGGGGGTGGTGTGCG	BamHI
	Rev	CCCGCTCGAG-TTGCTTGGCGGCAAGGC	XhoI
fu (961)-983-His	Fwd	CGCGGATCC - GGCGGAGGCGGCACCTT	BamHI
	Rev	CCCGCTCGAG-GAACCGGTAGCCTACG	XhoI
fu (961)- Orf46.1- His	Fwd	CGCGGATCCGGTGGTGGTGGT- TCAGATTGGCAAACGATT	BamHI
	Rev	CCCGCTCGAG-CGTATCATATTCACGTGC	XhoI
fu (961 c-L)- 741(MC58)	Fwd	CGCGGATCC -GGAGGGGGTGGTGTGCG	BamHI
	Rev	CCCGCTCGAG-TTATTGCTTGGCGGCAAG	XhoI
fu (961c-L)-983	Fwd	CGCGGATCC - GGCGGAGGCGGCACCTT	BamHI
	Rev	CCCGCTCGAG-TCAGAACCGGTAGCCTAC	XhoI
fu (961c-L)- Orf46.1	Fwd	CGCGGATCCGGTGGTGGTGGT- TCAGATTGGCAAACGATT	BamHI
	Rev	CCCGCTCGAG-TTACGTATCATATTCACGTGC	XhoI
961-(His/GST)	Fwd	CGCGGATCCCATATG-GCCACAAAGCGACGACG	BamHI-NdeI

(MC58)	Rev	CCCG <u>CTCGAG</u> -CCACTCGTAATTGACGCC	XhoI
961 Δ 1-His	Fwd	CGCGGAT <u>CCCATATG</u> -GCCACAAACGACGAC	NdeI
	Rev	CCCG <u>CTCGAG</u> -TGCTTGGCGGCAAAGTT	XhoI
961a-(His/GST)	Fwd	CGCGGAT <u>CCCATATG</u> -GCCACAAACGACGAC	BamHI-NdeI
	Rev	CCCG <u>CTCGAG</u> -TTAGCAATATTATCTTGTCTAGC	XhoI
961b-(His/GST)	Fwd	CGCGGAT <u>CCCATATG</u> -AAAGCAAACCGTGCAGA	BamHI-NdeI
	Rev	CCCG <u>CTCGAG</u> -CCACTCGTAATTGACGCC	XhoI
961-His/GST ^{GATE}	Fwd	ggggacaagttgtacaaaaaaagcaggctGCAGCCACAAACGACGACG ATGTTAAAAAAAGC	<i>attB1</i>
	Rev	ggggaccacttgcataagaaagctgggtTACCACTCGTAATTGACGC CGACATGGTAGG	<i>attB2</i>
982	Fwd	CGGG <u>CATATG</u> -GCAGCAAAAGACGTACAGTT	NdeI
	Rev	CGGG <u>CTCGAG</u> -TTACATCATGCCGCCATACCA	XhoI
983-His (2996)	Fwd	CGCGGAT <u>CCGCTAGC</u> -TTAGGCGGCGCGGAG	NheI
	Rev	CCCG <u>CTCGAG</u> -GAACCGGTAGCCTACG	XhoI
Δ 983-His (2996)	Fwd	CCC <u>CTAGCTAGC</u> -ACTTCTGCGCCCGACTT	NheI
	Rev	CCCG <u>CTCGAG</u> -GAACCGGTAGCCTACG	XhoI
983-His	Fwd	CGCGGAT <u>CCGCTAGC</u> -TTAGGCGGCGGGAG	NheI
	Rev	CCCG <u>CTCGAG</u> -GAACCGGTAGCCTACG	XhoI
Δ 983-His	Fwd	CGCGGAT <u>CCGCTAGC</u> -ACTTCTGCGCCCGACTT	NheI
	Rev	CCCG <u>CTCGAG</u> -GAACCGGTAGCCTACG	XhoI
983L	Fwd	CGCGGAT <u>CCGCTAGC</u> - CGAACGACCCAACCTCCCTACAAAAACTTCAA	NheI
	Rev	CCCG <u>CTCGAG</u> -TCAGAACCGACGTGCCAAGCCGTC	XhoI
987-His (MC58)	Fwd	GCCGCCATATGCCCCACTGGAAGAACGGACG	NdeI
	Rev	GCCGCCTCGAGTAATAAACCTTCTATGGCAGCAG	XhoI
989-(His/GST) (MC58)	Fwd	CGCGGAT <u>CCCATATG</u> -TCCGTCCACGCATCCG	BamHI-NdeI
	Rev	CCCG <u>CTCGAG</u> -TTGAATTGTAGGTGTATTG	XhoI
989L (MC58)	Fwd	CGCGGAT <u>CCCATATG</u> -ACCCCTCCGCACT	NdeI
	Rev	CCCG <u>CTCGAG</u> -TTATTGAATTGTAGGTGTAT	XhoI
CrgA-His (MC58)	Fwd	CGCGGAT <u>CCCATATG</u> -AAAACCAATTAGAAGAA	NdeI
	Rev	CCCG <u>CTCGAG</u> -TCCACAGAGATTGTTCC	XhoI
PilC1-ES (MC58)	Fwd	GATGCCCGAAGGGCGGG	
	Rev	GCCCA <u>AGCTT</u> -TCAGAACGACTTCACGC	
PilC1-His (MC58)	Fwd	CGCGGAT <u>CCCATATG</u> -CAAACCCATAAACACGCTATT	NdeI
	Rev	GCCCA <u>AGCTT</u> -GAAGAACGACTTCACGCCAG	HindIII
Δ 1PilC1-His (MC58)	Fwd	CGCGGAT <u>CCCATATG</u> -GTCTTTGACAATACCGA	NdeI
	Rev	GCCCA <u>AGCTT</u> -	HindIII
PilC1L (MC58)	Fwd	CGCGGAT <u>CCCATATG</u> -AATAAAACTTAAAAAGGCGG	NdeI
	Rev	GCCCA <u>AGCTT</u> -TCAGAACGACTTCACGC	HindIII
Δ GTbp2-His (MC58)	Fwd	CGCGAAT <u>CCCATATG</u> -TTCGATCTGATTCTGCGA	NdeI
	Rev	CCCG <u>CTCGAG</u> -TCGCACAGGCTGTTGGCG	XhoI
Tbp2-His (MC58)	Fwd	CGCGAAT <u>CCCATATG</u> -TTGGCGGAGGCAGCAG	NdeI
	Rev	CCCG <u>CTCGAG</u> -TCGCACAGGCTGTTGGCG	XhoI
Tbp2-His(MC58)	Fwd	CGCGAAT <u>CCCATATG</u> -TTGGCGGAGGCAGCAG	NdeI
	Rev	CCCG <u>CTCGAG</u> -TCGCACAGGCTGTTGGCG	XhoI

NMB0109- (His/GST) (MC58)	Fwd	CGCGGATCCCATATG-GCAAATTGGAGGTGCGC	BamHI-NdeI
	Rev	CCCGCTCGAG-TTCGGAGCGGTTGAAGC	XhoI
NMB0109L (MC58)	Fwd	CGCGGATCCCATATG-CAACGTCGTATTATAACCC	NdeI
	Rev	CCCGCTCGAG-TTATCGGAGCGGTTGAAG	XhoI
NMB0207- (His/GST) (MC58)	Fwd	CGCGGATCCCATATG- GGCATCAAAGTCGCCATCAACGGCTAC	BamHI-NdeI
	Rev	CCCGCTCGAG-TTGAGCGGGCGCACTCAAGTCCG	XhoI
NMB0462- (His/GST) (MC58)	Fwd	CGCGGATCCCATATG-GCGGGCAGCGAAAAAAAC	BamHI-NdeI
	Rev	CCCGCTCGAG-GTTGGTGCCGACTTTGAT	XhoI
NMB0623- (His/GST) (MC58)	Fwd	CGCGGATCCCATATG-GCGGGCGGAAGCGATA	BamHI-NdeI
	Rev	CCCGCTCGAG-TTGCCGCTTGAGCC	XhoI
NMB0625 (His- GST)(MC58)	Fwd	CGCGGATCCCATATGGCAAATCCGAAAATACG	BamHI-NdeI
	Rev	CCCGCTCGAGCATCCGTACTGTTCG	XhoI
NMB0634 (His/GST)(MC58)	Fwd	ggggacaagttgtacaaaaaagcaggctCCGACATTACCGTGTACAAC GCCAACAAAGAA	attB1
	Rev	ggggaccacttgtacaaaaagctgggtCTTATTCATACCGGCTTGCT CAAGCAGCCGG	attB2
NMB0776- His/GST (MC58) GATE	Fwd	ggggacaagttgtacaaaaaagcaggctGATACGGTGTTCCTGTAA AACGGACAAACAA	attB1
	Rev	ggggaccacttgtacaaaaagctgggtCTAGGAAAAATCGTCATCGT TGAAATTGCC	attB2
NMB1115- His/GST (MC58) GATE	Fwd	ggggacaagttgtacaaaaaagcaggctATGCACCCCATCGAAACC	attB1
	Rev	ggggaccacttgtacaaaaagctgggtCTAGTCTGCAGTGCCTC	attB2
NMB1343- (His/GST) (MC58)	Fwd	CGCGGATCCCATATG- GGAAATTCTTATATAGAGGCATTAG	BamHI-NdeI
	Rev	CCCGCTCGAG- GTTAATTCTATCAACTCTTAGCAATAAT	XhoI
NMB1369 (His- GST (MC58)	Fwd	CGCGGATCCCATATGGCCTGCCAGACGACA	BamHI-NdeI
	Rev	CCCGCTCGAGCCGCCTCTGCCAAA	XhoI
NMB1551 (His- GST)(MC58)	Fwd	CGCGGATCCCATATGGCAGAGATCTGTTGATAA	BamHI-NdeI
	Rev	CCCGCTCGAGCGGTTTCCGCCAATG	XhoI
NMB1899 (His- GST) (MC58)	Fwd	CGCGGATCCCATATG-CAGCCGGATACGGTC	BamHI-NdeI
	Rev	CCCGCTCGAGAATCACTCCAACACAAAAT	XhoI
NMB2050- (His/GST) (MC58)	Fwd	CGCGGATCCCATATG-TGGTTGCTGATGAAGGGC	BamHI-NdeI
	Rev	CCCGCTCGAG-GACTGCTTCATCTCTGC	XhoI
NMB2050L (MC58)	Fwd	CGCGGATCCCATATG-GAACTGATGACTGTTGC	NdeI
	Rev	CCCGCTCGAG-TCAGACTGCTTCATCTCT	XhoI
NMB2159- (His/GST) (MC58)	Fwd	CGCGGATCCCATATG- AGCATTAAAGTAGCGATTAACGGTTCGC	BamHI-NdeI
	Rev	CCCGCTCGAG- GATTTCGCTGCGAAGTATTCAAAGTGC	XhoI
fu-ΔG287....His	Fwd	CGCGGATCCGCTAGC-CCCGATGTTAAATCGGC	NheI

	Rev	CGGGGATCC-ATCCTGCTTTTTGCCGG	BamHI
fu-(ΔG287)-919-His	Fwd	CGCGGATCCGGTGGTGGTGGT-CAAAGCAAGAGCATCCAAACC	BamHI
	Rev	CCCAAGCTT-TTCGGCGGTATTCGGGCTTC	HindIII
fu-(ΔG287)-953-His	Fwd	CGCGGATCCGGTGGTGGTGGT-GCCACCTACAAAGTGGAC	BamHI
	Rev	GCCCAAGCTT-TTGTGCTGCCTCGAT	HindIII
fu-(ΔG287)-961-His	Fwd	CGCGGATCCGGTGGTGGTGGT-ACAAGCGACGACG	BamHI
	Rev	GCCCAAGCTT-CCACTCGTAATTGACGCC	HindIII
fu-(ΔG287)-Orf46.1-His	Fwd	CGCGGATCCGGTGGTGGTGGT-TCAGATTGGCAAACGATT	BamHI
	Rev	CCCAAGCTT-CGTATCATATTCACGTGC	HindIII
fu-(ΔG287-919)-Orf46.1-His	Fwd	CCCAAGCTTGGTGGTGGTGGTGGT-TCAGATTGGCAAACGATT	HindIII
	Rev	CCCGCTCGAG-CGTATCATATTCACGTGC	XhoI
fu-(ΔG287-Orf46.1)-919-His	Fwd	CCCAAGCTTGGTGGTGGTGGTGGT-CAAAGCAAGAGCATCCAAACC	HindIII
	Rev	CCCGCTCGAG-CGGCGGTATTCGGGCTT	XhoI
fu ΔG287(394.98)-...	Fwd	CGCGGATCCGCTAGC-CCCGATGTTAAATCGGC	NheI
	Rev	CGGGGATCC-ATCCTGCTTTTTGCCGG	BamHI
fu Orf1-(Orf46.1)-His	Fwd	CGCGGATCCGCTAGC-GGACACACTTATTCGGCATC	NheI
	Rev	CGCGGATCC-CCAGCGGTAGCCTAATTGAT	
fu (Orf1)-Orf46.1-His	Fwd	CGCGGATCCGGTGGTGGTGGT-TCAGATTGGCAAACGATT	BamHI
	Rev	CCCAAGCTT-CGTATCATATTCACGTGC	HindIII
fu (919)-Orf46.1-His	Fwd1	GCGCGTCGACGGTGGCGGAGGCACTGGATCCTCAG	SalII
	Fwd2	GGAGGCACTGGATCCTCAGATTGGCAAACGATT	
	Rev	CCCGCTCGAG-CGTATCATATTCACGTGC	XhoI
Fu orf46-....	Fwd	GGAATTCCATATGTCAGATTGGCAAACGATT	NdeI
	Rev	CGCGGATCCCGTATCATATTCACGTGC	BamHI
Fu (orf46)-287-His	Fwd	CGGGGATCCGGGGCGGCGGTGGCG	BamHI
	Rev	CCCAAGCTTATCCTGCTTTTTGCCGGC	HindIII
Fu (orf46)-919-His	Fwd	CGCGGATCCGGTGGTGGTCAAAGCAAGAGCATCCA AACC	BamHI
	Rev	CCCAAGCTTGGCGGTATTCGGGCTTC	HindIII
Fu (orf46-919)-287-His	Fwd	CCCCAAGCTTGGGGCGGCGGTGGCG	HindIII
	Rev	CCCGCTCGAGATCCTGCTCTTTTGCCGGC	XhoI
Fu (orf46-287)-919-His	Fwd	CCCAAGCTTGGTGGTGGTGGTCAAAGCAAGAGCAT CCAAACC	HindIII
	Rev	CCCGCTCGAGCGGGCGGTATTCGGGCTT	XhoI
(ΔG741)-961c-His	Fwd1	GGAGGCACTGGATCCGCAGCCACAAACGACGACGA	XhoI
	Fwd2	GCGGCCTCGAG-GGTGGCGGAGGCACTGGATCCGCAG	
	Rev	CCCGCTCGAG-ACCCAGCTTGTAAAGGTG	XhoI
(ΔG741)-961-His	Fwd1	GGAGGCACTGGATCCGCAGCCACAAACGACGACGA	XhoI
	Fwd2	GCGGCCTCGAG-GGTGGCGGAGGCACTGGATCCGCAG	
	Rev	CCCGCTCGAG-CCACTCGTAATTGACGCC	XhoI

(Δ G741)-983-His	Fwd	GGGGC <u>CTCGAG</u> - GGATCCGGCGGAGGC GG CACTTCTGCG	XhoI
	Rev	CCCG <u>CTCGAG</u> -GAACCGGTAGCCTACG	XhoI
(Δ G741)-orf46.1-His	Fwd1	GGAGGC <u>ACTGGATCCTCAGATTGGCAAACGATTC</u>	SalI
	Fwd2	GGGG <u>CGTCGACGGTGGCGGAGGC</u> ACTGGATCCTCAGA	
	Rev	CCCG <u>CTCGAG</u> -CGTATCATATTCACGTGC	XhoI
(Δ G983)-741(MC58)-His	Fwd	GGGG <u>CTCGAG</u> -GGATCCGGAGGGGGTGGTGTGCC	XhoI
	Rev	CCCG <u>CTCGAG</u> -TTGCTTGGCGGCAAG	XhoI
(Δ G983)-961c-His	Fwd1	GGAGGC <u>ACTGGATCCGCAGCCACAAACGACGACGA</u>	XhoI
	Fwd2	GGGG <u>CTCGAG</u> -GGTGGCGGAGGC <u>ACTGGATCCGCAG</u>	
	Rev	CCCG <u>CTCGAG</u> -ACCCAGCTTGTAAAGTTG	XhoI
(Δ G983)-961-His	Fwd1	GGAGGC <u>ACTGGATCCGCAGCCACAAACGACGACGA</u>	XhoI
	Fwd2	GGGG <u>CTCGAG</u> -GGTGGCGGAGGC <u>ACTGGATCCGCAG</u>	
	Rev	CCCG <u>CTCGAG</u> -CCACTCGTAATTGACGCC	XhoI
(Δ G983)-Orf46.1-His	Fwd1	GGAGGC <u>ACTGGATCCTCAGATTGGCAAACGATTC</u>	SalI
	Fwd2	GGGG <u>CGTCGACGGTGGCGGAGGC</u> ACTGGATCCTCAGA	
	Rev	CCCG <u>CTCGAG</u> -CGTATCATATTCACGTGC	XhoI

* This primer was used as a Reverse primer for all the C terminal fusions of 287 to the His-tag.

[§] Forward primers used in combination with the 287-His Reverse primer.

NB – All PCR reactions use strain 2996 unless otherwise specified (e.g. strain MC58)

In all constructs starting with an ATG not followed by a unique *NheI* site, the ATG codon is
5 part of the *NdeI* site used for cloning. The constructs made using *NheI* as a cloning site at the
5' end (e.g. all those containing 287 at the N-terminus) have two additional codons (GCT
AGC) fused to the coding sequence of the antigen.

Preparation of chromosomal DNA templates

N.meningitidis strains 2996, MC58, 394.98, 1000 and BZ232 (and others) were grown to
10 exponential phase in 100ml of GC medium, harvested by centrifugation, and resuspended in
5ml buffer (20% w/v sucrose, 50mM Tris-HCl, 50mM EDTA, pH8). After 10 minutes
incubation on ice, the bacteria were lysed by adding 10ml of lysis solution (50mM NaCl, 1%
Na-Sarkosyl, 50 μ g/ml Proteinase K), and the suspension incubated at 37°C for 2 hours. Two
15 phenol extractions (equilibrated to pH 8) and one CHCl₃/isoamylalcohol (24:1) extraction
were performed. DNA was precipitated by addition of 0.3M sodium acetate and 2 volumes
of ethanol, and collected by centrifugation. The pellet was washed once with 70%(v/v)
ethanol and redissolved in 4.0ml TE buffer (10mM Tris-HCl, 1mM EDTA, pH 8.0). The
DNA concentration was measured by reading OD₂₆₀.

PCR Amplification

The standard PCR protocol was as follows: 200ng of genomic DNA from 2996, MC581000, or BZ232 strains or 10ng of plasmid DNA preparation of recombinant clones were used as template in the presence of 40 μ M of each oligonucleotide primer, 400-800 μ M dNTPs
5 solution, 1x PCR buffer (including 1.5mM MgCl₂), 2.5 units *TaqI* DNA polymerase (using Perkin-Elmer AmpliTaq, Boerhinger Mannheim ExpandTM Long Template).

After a preliminary 3 minute incubation of the whole mix at 95°C, each sample underwent a two-step amplification: the first 5 cycles were performed using the hybridisation temperature that excluded the restriction enzyme tail of the primer (T_{m1}). This was followed by 30 cycles
10 according to the hybridisation temperature calculated for the whole length oligos (T_{m2}). Elongation times, performed at 68°C or 72°C, varied according to the length of the Orf to be amplified. In the case of Orf1 the elongation time, starting from 3 minutes, was increased by 15 seconds each cycle. The cycles were completed with a 10 minute extension step at 72°C.

The amplified DNA was either loaded directly on a 1% agarose gel. The DNA fragment
15 corresponding to the band of correct size was purified from the gel using the Qiagen Gel Extraction Kit, following the manufacturer's protocol.

Digestion of PCR fragments and of the cloning vectors

The purified DNA corresponding to the amplified fragment was digested with the appropriate restriction enzymes for cloning into pET-21b+, pET22b+ or pET-24b+. Digested
20 fragments were purified using the QIAquick PCR purification kit (following the manufacturer's instructions) and eluted with either H₂O or 10mM Tris, pH 8.5. Plasmid vectors were digested with the appropriate restriction enzymes, loaded onto a 1.0% agarose gel and the band corresponding to the digested vector purified using the Qiagen QIAquick Gel Extraction Kit.

25 *Cloning*

The fragments corresponding to each gene, previously digested and purified, were ligated into pET21b+, pET22b+ or pET-24b+. A molar ratio of 3:1 fragment/vector was used with T4 DNA ligase in the ligation buffer supplied by the manufacturer.

Recombinant plasmid was transformed into competent *E.coli* DH5 or HB101 by incubating
30 the ligase reaction solution and bacteria for 40 minutes on ice, then at 37°C for 3 minutes.

This was followed by the addition of 800 μ l LB broth and incubation at 37°C for 20 minutes. The cells were centrifuged at maximum speed in an Eppendorf microfuge, resuspended in approximately 200 μ l of the supernatant and plated onto LB ampicillin (100mg/ml) agar.

Screening for recombinant clones was performed by growing randomly selected colonies 5 overnight at 37°C in 4.0ml of LB broth + 100 μ g/ml ampicillin. Cells were pelleted and plasmid DNA extracted using the Qiagen QIAprep Spin Miniprep Kit, following the manufacturer's instructions. Approximately 1 μ g of each individual miniprep was digested with the appropriate restriction enzymes and the digest loaded onto a 1-1.5% agarose gel (depending on the expected insert size), in parallel with the molecular weight marker (1kb 10 DNA Ladder, GIBCO). Positive clones were selected on the basis of the size of insert.

Expression

After cloning each gene into the expression vector, recombinant plasmids were transformed into *E.coli* strains suitable for expression of the recombinant protein. 1 μ l of each construct was used to transform *E.coli* BL21-DE3 as described above. Single recombinant colonies 15 were inoculated into 2ml LB+Amp (100 μ g/ml), incubated at 37°C overnight, then diluted 1:30 in 20ml of LB+Amp (100 μ g/ml) in 100ml flasks, to give an OD₆₀₀ between 0.1 and 0.2. The flasks were incubated at 30°C or at 37°C in a gyratory water bath shaker until OD₆₀₀ indicated exponential growth suitable for induction of expression (0.4-0.8 OD). Protein expression was induced by addition of 1.0mM IPTG. After 3 hours incubation at 30°C or 20 37°C the OD₆₀₀ was measured and expression examined. 1.0ml of each sample was centrifuged in a microfuge, the pellet resuspended in PBS and analysed by SDS-PAGE and Coomassie Blue staining.

Gateway cloning and expression

Sequences labelled GATE were cloned and expressed using the GATEWAY Cloning 25 Technology (GIBCO-BRL). Recombinational cloning (RC) is based on the recombination reactions that mediate the integration and excision of phage into and from the *E.coli* genome, respectively. The integration involves recombination of the *attP* site of the phage DNA within the *attB* site located in the bacterial genome (BP reaction) and generates an integrated phage genome flanked by *attL* and *attR* sites. The excision recombines *attL* and *attR* sites back to *attP* 30 and *attB* sites (LR reaction). The integration reaction requires two enzymes [the phage protein Integrase (Int) and the bacterial protein integration host factor (IHF)] (BP clonase). The

excision reaction requires Int, IHF, and an additional phage enzyme, Excisionase (Xis) (LR clonase). Artificial derivatives of the 25-bp bacterial *attB* recombination site, referred to as B1 and B2, were added to the 5' end of the primers used in PCR reactions to amplify Neisserial ORFs. The resulting products were BP cloned into a "Donor vector" containing complementary 5 derivatives of the phage *attP* recombination site (P1 and P2) using BP clonase. The resulting "Entry clones" contain ORFs flanked by derivatives of the *attL* site (L1 and L2) and were subcloned into expression "destination vectors" which contain derivatives of the *attL*-compatible *attR* sites (R1 and R2) using LR clonase. This resulted in "expression clones" in which ORFs are flanked by B1 and B2 and fused in frame to the GST or His N terminal tags.

10 The *E. coli* strain used for GATEWAY expression is BL21-SI. Cells of this strain are induced for expression of the T7 RNA polymerase by growth in medium containing salt (0.3 M NaCl).

Note that this system gives N-terminus His tags.

Preparation of membrane proteins.

Fractions composed principally of either inner, outer or total membrane were isolated in 15 order to obtain recombinant proteins expressed with membrane-localisation leader sequences. The method for preparation of membrane fractions, enriched for recombinant proteins, was adapted from Filip *et. al.* [J.Bact. (1973) 115:717-722] and Davies *et. al.* [J.Immunol.Meth. (1990) 143:215-225]. Single colonies harbouring the plasmid of interest 20 were grown overnight at 37°C in 20 ml of LB/Amp (100 µg/ml) liquid culture. Bacteria were diluted 1:30 in 1.0 L of fresh medium and grown at either 30°C or 37°C until the OD₅₅₀ 25 reached 0.6-0.8. Expression of recombinant protein was induced with IPTG at a final concentration of 1.0 mM. After incubation for 3 hours, bacteria were harvested by centrifugation at 8000g for 15 minutes at 4°C and resuspended in 20 ml of 20 mM Tris-HCl (pH 7.5) and complete protease inhibitors (Boehringer-Mannheim). All subsequent procedures were performed at 4°C or on ice.

Cells were disrupted by sonication using a Branson Sonifier 450 and centrifuged at 5000g for 20 min to sediment unbroken cells and inclusion bodies. The supernatant, containing membranes and cellular debris, was centrifuged at 50000g (Beckman Ti50, 29000rpm) for 75 min, washed with 20 mM Bis-tris propane (pH 6.5), 1.0 M NaCl, 10% (v/v) glycerol and 30 sedimented again at 50000g for 75 minutes. The pellet was resuspended in 20mM Tris-HCl (pH 7.5), 2.0% (v/v) Sarkosyl, complete protease inhibitor (1.0 mM EDTA, final

concentration) and incubated for 20 minutes to dissolve inner membrane. Cellular debris was pelleted by centrifugation at 5000g for 10 min and the supernatant centrifuged at 75000g for 75 minutes (Beckman Ti50, 33000rpm). Proteins 008L and 519L were found in the supernatant suggesting inner membrane localisation. For these proteins both inner and total membrane fractions (washed with NaCl as above) were used to immunise mice. Outer membrane vesicles obtained from the 75000g pellet were washed with 20 mM Tris-HCl (pH 7.5) and centrifuged at 75000g for 75 minutes or overnight. The OMV was finally resuspended in 500 µl of 20 mM Tris-HCl (pH 7.5), 10% v/v glycerol. Orf1L and Orf40L were both localised and enriched in the outer membrane fraction which was used to immunise mice. Protein concentration was estimated by standard Bradford Assay (Bio-Rad), while protein concentration of inner membrane fraction was determined with the DC protein assay (Bio-Rad). Various fractions from the isolation procedure were assayed by SDS-PAGE.

Purification of His-tagged proteins

Various forms of 287 were cloned from strains 2996 and MC58. They were constructed with a C-terminus His-tagged fusion and included a mature form (aa 18-427), constructs with deletions (Δ 1, Δ 2, Δ 3 and Δ 4) and clones composed of either B or C domains. For each clone purified as a His-fusion, a single colony was streaked and grown overnight at 37°C on a LB/Amp (100 µg/ml) agar plate. An isolated colony from this plate was inoculated into 20ml of LB/Amp (100 µg/ml) liquid medium and grown overnight at 37°C with shaking. The overnight culture was diluted 1:30 into 1.0 L LB/Amp (100 µg/ml) liquid medium and allowed to grow at the optimal temperature (30 or 37°C) until the OD₅₅₀ reached 0.6-0.8. Expression of recombinant protein was induced by addition of IPTG (final concentration 1.0mM) and the culture incubated for a further 3 hours. Bacteria were harvested by centrifugation at 8000g for 15 min at 4°C. The bacterial pellet was resuspended in 7.5 ml of either (i) cold buffer A (300 mM NaCl, 50 mM phosphate buffer, 10 mM imidazole, pH 8.0) for soluble proteins or (ii) buffer B (10mM Tris-HCl, 100 mM phosphate buffer, pH 8.8 and, optionally, 8M urea) for insoluble proteins. Proteins purified in a soluble form included 287-His, Δ 1, Δ 2, Δ 3 and Δ 4287-His, Δ 4287MC58-His, 287c-His and 287cMC58-His. Protein 287bMC58-His was insoluble and purified accordingly. Cells were disrupted by sonication on ice four times for 30 sec at 40W using a Branson sonifier 450 and centrifuged at 13000xg for 30 min at 4°C. For insoluble proteins, pellets were resuspended in 2.0 ml buffer C (6 M guanidine hydrochloride, 100 mM phosphate buffer, 10 mM Tris- HCl, pH 7.5

and treated with 10 passes of a Dounce homogenizer. The homogenate was centrifuged at 13000g for 30 min and the supernatant retained. Supernatants for both soluble and insoluble preparations were mixed with 150µl Ni²⁺-resin (previously equilibrated with either buffer A or buffer B, as appropriate) and incubated at room temperature with gentle agitation for 30 5 min. The resin was Chelating Sepharose Fast Flow (Pharmacia), prepared according to the manufacturer's protocol. The batch-wise preparation was centrifuged at 700g for 5 min at 4°C and the supernatant discarded. The resin was washed twice (batch-wise) with 10ml buffer A or B for 10 min, resuspended in 1.0 ml buffer A or B and loaded onto a disposable 10 column. The resin continued to be washed with either (i) buffer A at 4°C or (ii) buffer B at room temperature, until the OD₂₈₀ of the flow-through reached 0.02-0.01. The resin was further washed with either (i) cold buffer C (300mM NaCl, 50mM phosphate buffer, 20mM imidazole, pH 8.0) or (ii) buffer D (10mM Tris-HCl, 100mM phosphate buffer, pH 6.3 and, optionally, 8M urea) until OD₂₈₀ of the flow-through reached 0.02-0.01. The His-fusion 15 protein was eluted by addition of 700µl of either (i) cold elution buffer A (300 mM NaCl, 50mM phosphate buffer, 250 mM imidazole, pH 8.0) or (ii) elution buffer B (10 mM Tris-HCl, 100 mM phosphate buffer, pH 4.5 and, optionally, 8M urea) and fractions collected until the OD₂₈₀ indicated all the recombinant protein was obtained. 20µl aliquots of each elution fraction were analysed by SDS-PAGE. Protein concentrations were estimated using the Bradford assay.

20 ***Renaturation of denatured His-fusion proteins.***

Denaturation was required to solubilize 287bMC8, so a renaturation step was employed prior to immunisation. Glycerol was added to the denatured fractions obtained above to give a final concentration of 10% v/v. The proteins were diluted to 200 µg/ml using dialysis buffer I (10% v/v glycerol, 0.5M arginine, 50 mM phosphate buffer, 5.0 mM reduced glutathione, 25 0.5 mM oxidised glutathione, 2.0M urea, pH 8.8) and dialysed against the same buffer for 12-14 hours at 4°C. Further dialysis was performed with buffer II (10% v/v glycerol, 0.5M arginine, 50mM phosphate buffer, 5.0mM reduced glutathione, 0.5mM oxidised glutathione, pH 8.8) for 12-14 hours at 4°C. Protein concentration was estimated using the formula:

$$\text{Protein (mg/ml)} = (1.55 \times OD_{280}) - (0.76 \times OD_{260})$$

Amino acid sequence analysis.

Automated sequence analysis of the NH₂-terminus of proteins was performed on a Beckman sequencer (LF 3000) equipped with an on-line phenylthiohydantoin-amino acid analyser (System Gold) according to the manufacturer's recommendations.

5 *Immunization*

Balb/C mice were immunized with antigens on days 0, 21 and 35 and sera analyzed at day 49.

Sera analysis – ELISA

The acapsulated MenB M7 and the capsulated strains were plated on chocolate agar plates and incubated overnight at 37°C with 5% CO₂. Bacterial colonies were collected from the

10 agar plates using a sterile dracon swab and inoculated into Mueller-Hinton Broth (Difco) containing 0.25% glucose. Bacterial growth was monitored every 30 minutes by following OD₆₂₀. The bacteria were let to grow until the OD reached the value of 0.4-0.5. The culture was centrifuged for 10 minutes at 4000rpm. The supernatant was discarded and bacteria were washed twice with PBS, resuspended in PBS containing 0.025% formaldehyde, and
15 incubated for 1 hour at 37°C and then overnight at 4°C with stirring. 100µl bacterial cells were added to each well of a 96 well Greiner plate and incubated overnight at 4°C. The wells were then washed three times with PBT washing buffer (0.1% Tween-20 in PBS). 200µl of saturation buffer (2.7% polyvinylpyrrolidone 10 in water) was added to each well and the plates incubated for 2 hours at 37°C. Wells were washed three times with PBT. 200µl of
20 diluted sera (Dilution buffer: 1% BSA, 0.1% Tween-20, 0.1% NaN₃ in PBS) were added to each well and the plates incubated for 2 hours at 37°C. Wells were washed three times with PBT. 100µl of HRP-conjugated rabbit anti-mouse (Dako) serum diluted 1:2000 in dilution buffer were added to each well and the plates were incubated for 90 minutes at 37°C. Wells were washed three times with PBT buffer. 100µl of substrate buffer for HRP (25ml of citrate
25 buffer pH5, 10mg of O-phenildiamine and 10µl of H₂O₂) were added to each well and the plates were left at room temperature for 20 minutes. 100µl 12.5% H₂SO₄ was added to each well and OD₄₉₀ was followed. The ELISA titers were calculated arbitrarily as the dilution of sera which gave an OD₄₉₀ value of 0.4 above the level of preimmune sera. The ELISA was considered positive when the dilution of sera with OD₄₉₀ of 0.4 was higher than 1:400.

30 *Sera analysis – FACS Scan bacteria binding assay*

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C with 5% CO₂. Bacterial colonies were collected from the agar plates using

a sterile dracon swab and inoculated into 4 tubes containing 8ml each Mueller-Hinton Broth (Difco) containing 0.25% glucose. Bacterial growth was monitored every 30 minutes by following OD₆₂₀. The bacteria were let to grow until the OD reached the value of 0.35-0.5. The culture was centrifuged for 10 minutes at 4000rpm. The supernatant was discarded and

5 the pellet was resuspended in blocking buffer (1% BSA in PBS, 0.4% NaN₃) and centrifuged for 5 minutes at 4000rpm. Cells were resuspended in blocking buffer to reach OD₆₂₀ of 0.05. 100μl bacterial cells were added to each well of a Costar 96 well plate. 100μl of diluted (1:100, 1:200, 1:400) sera (in blocking buffer) were added to each well and plates incubated for 2 hours at 4°C. Cells were centrifuged for 5 minutes at 4000rpm, the supernatant

10 aspirated and cells washed by addition of 200μl/well of blocking buffer in each well. 100μl of R-Phicoerytrin conjugated F(ab)₂ goat anti-mouse, diluted 1:100, was added to each well and plates incubated for 1 hour at 4°C. Cells were spun down by centrifugation at 4000rpm for 5 minutes and washed by addition of 200μl/well of blocking buffer. The supernatant was aspirated and cells resuspended in 200μl/well of PBS, 0.25% formaldehyde. Samples were

15 transferred to FACScan tubes and read. The condition for FACScan (Laser Power 15mW) setting were: FL2 on; FSC-H threshold:92; FSC PMT Voltage: E 01; SSC PMT: 474; Amp. Gains 6.1; FL-2 PMT: 586; compensation values: 0.

Sera analysis – bactericidal assay

N. meningitidis strain 2996 was grown overnight at 37°C on chocolate agar plates (starting 20 from a frozen stock) with 5% CO₂. Colonies were collected and used to inoculate 7ml Mueller-Hinton broth, containing 0.25% glucose to reach an OD₆₂₀ of 0.05-0.08. The culture was incubated for approximately 1.5 hours at 37 degrees with shacking until the OD₆₂₀ reached the value of 0.23-0.24. Bacteria were diluted in 50mM Phosphate buffer pH 7.2 containing 10mM MgCl₂, 10mM CaCl₂ and 0.5% (w/v) BSA (assay buffer) at the working 25 dilution of 10⁵ CFU/ml. The total volume of the final reaction mixture was 50 μl with 25 μl of serial two fold dilution of test serum, 12.5 μl of bacteria at the working dilution, 12.5 μl of baby rabbit complement (final concentration 25%).

Controls included bacteria incubated with complement serum, immune sera incubated with 30 bacteria and with complement inactivated by heating at 56°C for 30'. Immediately after the addition of the baby rabbit complement, 10μl of the controls were plated on Mueller-Hinton agar plates using the tilt method (time 0). The 96-wells plate was incubated for 1 hour at 37°C with rotation. 7μl of each sample were plated on Mueller-Hinton agar plates as spots, whereas 10μl of the controls were plated on Mueller-Hinton agar plates using the tilt method

-100-

(time 1). Agar plates were incubated for 18 hours at 37 degrees and the colonies corresponding to time 0 and time 1 were counted.

Sera analysis – western blots

Purified proteins (500ng/lane), outer membrane vesicles (5 μ g) and total cell extracts (25 μ g) derived from MenB strain 2996 were loaded onto a 12% SDS-polyacrylamide gel and transferred to a nitrocellulose membrane. The transfer was performed for 2 hours at 150mA at 4°C, using transfer buffer (0.3% Tris base, 1.44% glycine, 20% (v/v) methanol). The membrane was saturated by overnight incubation at 4°C in saturation buffer (10% skimmed milk, 0.1% Triton X100 in PBS). The membrane was washed twice with washing buffer (3% skimmed milk, 0.1% Triton X100 in PBS) and incubated for 2 hours at 37°C with mice sera diluted 1:200 in washing buffer. The membrane was washed twice and incubated for 90 minutes with a 1:2000 dilution of horseradish peroxidase labelled anti-mouse Ig. The membrane was washed twice with 0.1% Triton X100 in PBS and developed with the Opti-4CN Substrate Kit (Bio-Rad). The reaction was stopped by adding water.

The OMVs were prepared as follows: *N. meningitidis* strain 2996 was grown overnight at 37 degrees with 5% CO₂ on 5 GC plates, harvested with a loop and resuspended in 10 ml of 20mM Tris-HCl pH 7.5, 2 mM EDTA. Heat inactivation was performed at 56°C for 45 minutes and the bacteria disrupted by sonication for 5 minutes on ice (50% duty cycle, 50% output, Branson sonifier 3 mm microtip). Unbroken cells were removed by centrifugation at 5000g for 10 minutes, the supernatant containing the total cell envelope fraction recovered and further centrifuged overnight at 50000g at the temperature of 4°C. The pellet containing the membranes was resuspended in 2% sarkosyl, 20mM Tris-HCl pH 7.5, 2 mM EDTA and incubated at room temperature for 20 minutes to solubilise the inner membranes. The suspension was centrifuged at 10000g for 10 minutes to remove aggregates, the supernatant was further centrifuged at 50000g for 3 hours. The pellet, containing the outer membranes was washed in PBS and resuspended in the same buffer. Protein concentration was measured by the D.C. Bio-Rad Protein assay (Modified Lowry method), using BSA as a standard.

Total cell extracts were prepared as follows: *N. meningitidis* strain 2996 was grown overnight on a GC plate, harvested with a loop and resuspended in 1ml of 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes.

961 domain studies

Cellular fractions preparation Total lysate, periplasm, supernatant and OMV of *E.coli* clones expressing different domains of 961 were prepared using bacteria from over-night cultures or

after 3 hours induction with IPTG. Briefly, the periplasm were obtained suspending bacteria in saccarose 25% and Tris 50mM (pH 8) with polimixine 100 μ g/ml. After 1hr at room temperature bacteria were centrifuged at 13000rpm for 15 min and the supernatant were collected. The culture supernatant were filtered with 0.2 μ m and precipitated with TCA 50%

5 in ice for two hours. After centrifugation (30 min at 13000 rp) pellets were rinsed twice with ethanol 70% and suspended in PBS. The OMV preparation was performed as previously described. Each cellular fraction were analyzed in SDS-PAGE or in Western Blot using the polyclonal anti-serum raised against GST-961.

10 Adhesion assay Chang epithelial cells (Wong-Kilbourne derivative, clone 1-5c-4, human conjunctiva) were maintained in DMEM (Gibco) supplemented with 10% heat-inactivated FCS, 15mM L-glutamine and antibiotics.

For the adherence assay, sub-confluent culture of Chang epithelial cells were rinsed with PBS and treated with trypsin-EDTA (Gibco), to release them from the plastic support. The cells were then suspended in PBS, counted and dilute in PBS to 5×10^5 cells/ml.

15 Bacteria from over-night cultures or after induction with IPTG, were pelleted and washed twice with PBS by centrifuging at 13000 for 5 min. Approximately $2-3 \times 10^8$ (cfu) were incubated with 0.5 mg/ml FITC (Sigma) in 1ml buffer containing 50mM NaHCO₃ and 100mM NaCl pH 8, for 30 min at room temperature in the dark. FITC-labeled bacteria were wash 2-3 times and suspended in PBS at $1-1.5 \times 10^9$ /ml. 200 μ l of this suspension ($2-3 \times 10^8$)
20 were incubated with 200 μ l (1×10^5) epithelial cells for 30min a 37°C. Cells were than centrifuged at 2000rpm for 5 min to remove non-adherent bacteria, suspended in 200 μ l of PBS, transferred to FACScan tubes and read

CLAIMS

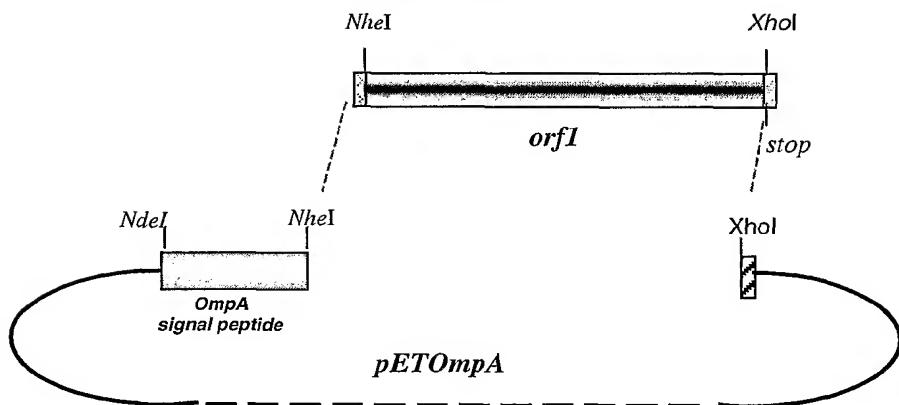
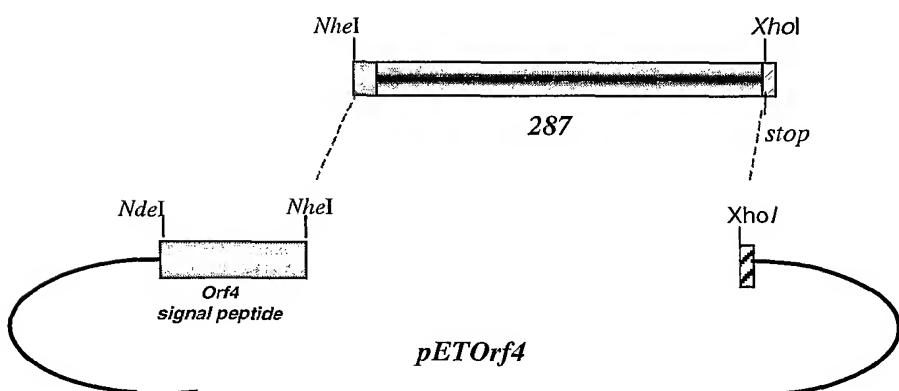
1. A method for the heterologous expression of a protein of the invention, in which (a) at least one domain in the protein is deleted and, optionally, (b) no fusion partner is used.
2. The method of claim 1, in which the protein of the invention is ORF46.
- 5 3. The method of claim 2, in which ORF46 is divided into a first domain (amino acids 1-433) and a second domain (amino acids 433-608).
4. The method of claim 2, in which the protein of the invention is 564.
5. The method of claim 4, in which protein 564 is divided into domains as shown in Figure 8.
- 10 6. The method of claim 1 in which the protein of the invention is 961.
7. The method of claim 6, in which protein 961 is divided into domains as shown in Figure 12.
8. The method of claim 1, in which the protein of the invention is 502 and the domain is amino acids 28 to 167 (numbered according to the MC58 sequence).
- 15 9. The method of claim 1, in which the protein of the invention is 287.
10. A method for the heterologous expression of a protein of the invention, in which (a) a portion of the N-terminal domain of the protein is deleted.
11. The method of claim 9 or claim 10, in which protein 287 is divided into domains A B & C shown in Figure 5.
- 20 12. The method of claim 11, in which (i) domain A, (ii) domains A and B, or (iii) domains A and C are deleted.
13. The method of claim 11, wherein (i) amino acids 1-17, (ii) amino acids 1-25, (iii) amino acids 1-69, or (iv) amino acids 1-106, of domain A are deleted.
14. A method for the heterologous expression of a protein of the invention, in which (a) no 25 fusion partner is used, and (b) the protein's native leader peptide (if present) is used.

15. The method of claim 14, in which the protein of the invention is selected from the group consisting of: 111, 149, 206, 225-1, 235, 247-1, 274, 283, 286, 292, 401, 406, 502-1, 503, 519-1, 525-1, 552, 556, 557, 570, 576-1, 580, 583, 664, 759, 907, 913, 920-1, 936-1, 953, 961, 983, 989, Orf4, Orf7-1, Orf9-1, Orf23, Orf25, Orf37, Orf38, Orf40, Orf40.1, 5 Orf40.2, Orf72-1, Orf76-1, Orf85-2, Orf91, Orf97-1, Orf119, Orf143.1, NMB0109, NMB2050, 008, 105, 117-1, 121-1, 122-1, 128-1, 148, 216, 243, 308, 593, 652, 726, 926, 982, Orf83-1 and Orf143-1.
16. A method for the heterologous expression of a protein of the invention, in which (a) the protein's leader peptide is replaced by the leader peptide from a different protein and, 10 optionally, (b) no fusion partner is used.
17. The method of claim 16, in which the different protein is 961, ORF4, *E.coli* OmpA, or *E.carotovora* PelB, or in which the leader peptide is MKKYLFSAA.
18. The method of claim 17, in which the different protein is *E.coli* OmpA and the protein of the invention is ORF1.
- 15 19. The method of claim 17, in which the protein of the invention is 911 and the different protein is *E.carotovora* PelB or *E.coli* OmpA.
20. The method of claim 17, in which the different protein is ORF4 and the protein of the invention is 287.
21. A method for the heterologous expression of a protein of the invention, in which (a) the protein's leader peptide is deleted and, optionally, (b) no fusion partner is used. 20
22. The method of claim 21, in which the protein of the invention is 919.
23. A method for the heterologous expression of a protein of the invention, in which expression of a protein of the invention is carried out at a temperature at which a toxic activity of the protein is not manifested.
- 25 24. The method of claim 23, in which protein 919 is expressed at 30°C.
25. A method for the heterologous expression of a protein of the invention, in which protein is mutated to reduce or eliminate toxic activity.

26. The method of claim 25, in which the protein of the invention is 907, 919 or 922.
27. The method of claim 26, in which 907 is mutated at Glu-117 (e.g. Glu→Gly).
28. The method of claim 26, in which 919 is mutated at Glu-255 (e.g. Glu→Gly) and/or Glu-323 (e.g. Glu→Gly).
- 5 29. The method of claim 26, in which 922 is mutated at Glu-164 (e.g. Glu→Gly), Ser-213 (e.g. Ser→Gly) and/or Asn-348 (e.g. Asn→Gly).
30. A method for the heterologous expression of a protein of the invention, in which vector pSM214 is used or vector pET-24b is used.
- 10 31. The method of claim 30, in which the protein of the invention is 953 and the vector is pSM214.
32. A method for the heterologous expression of a protein of the invention, in which a protein is expressed or purified such that it adopts a particular multimeric form.
33. The method of claim 32, in which protein 953 is expressed and/or purified in monomeric form.
- 15 34. The method of claim 32, in which protein 961 is expressed and/or purified in tetrameric form.
35. The method of claim 32, in which protein 287 is expressed and/or purified in dimeric form.
36. The method of claim 32, in which protein 919 is expressed and/or purified in monomeric 20 form.
37. A method for the heterologous expression of a protein of the invention, in which the protein is expressed as a lipidated protein.
38. The method of claim 37, in which the protein of the invention is 919, 287, ORF4, 406, 576, or ORF25.
- 25 39. A method for the heterologous expression of a protein of the invention, in which (a) the protein's C-terminus region is mutated and, optionally, (b) no fusion partner is used.

40. The method of claim 39, wherein the mutation is a substitution, an insertion, or a deletion
41. The method of claim 40, wherein the protein of the invention is 730, ORF29 or ORF46.
42. A method for the heterologous expression of a protein of the invention, in which the protein's leader peptide is mutated.
- 5 43. The method of claim 42, in which the protein of the invention is 919.
44. A method for the heterologous expression of a protein, in which a poly-glycine stretch within the protein is mutated.
45. The method of claim 44, wherein the protein is a protein of the invention.
46. The method of claim 45, wherein the protein of the invention is 287, 741, 983 or Tbp2.
- 10 47. The method of claim 46, wherein (Gly)₆ is deleted from 287 or 983.
48. The method of claim 46, wherein (Gly)₄ is deleted from Tbp2 or 741
49. The method of claim 47 or claim 48, wherein the leader peptide is also deleted.
50. The method of any preceding claim, in which the heterologous expression is in an *E.coli* host.
- 15 51. A protein expressed by the method of any preceding claim.
52. A heterologous protein comprising the N-terminal amino acid sequence MKKYLFSAA.

1/13

FIGURE 1**FIGURE 2**

2/13

FIGURE 3

M1 ORF1

**PURIFICATION**

S TP

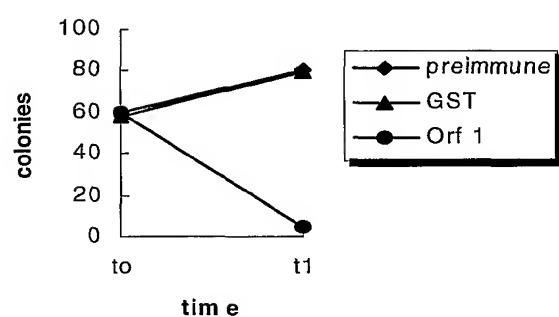
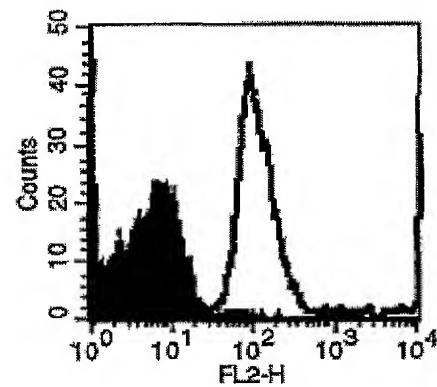
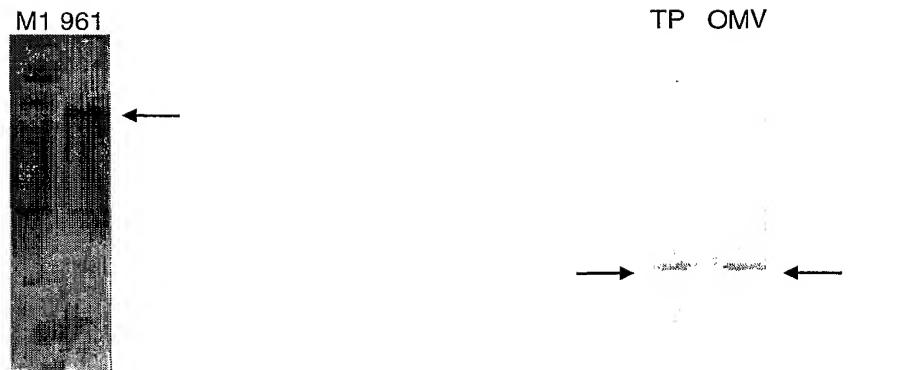
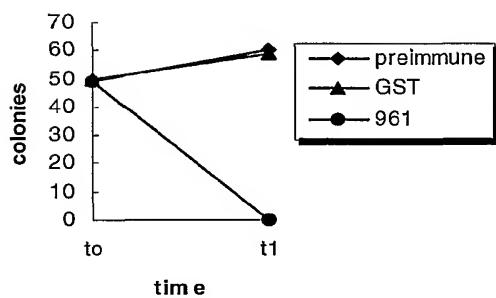
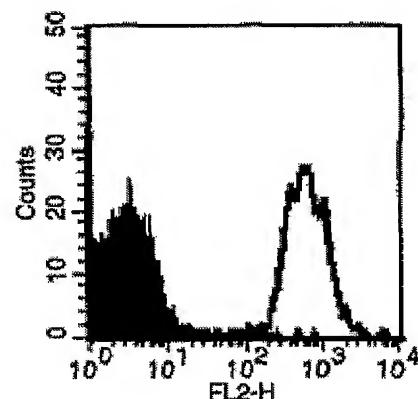
**WESTERN BLOTH****BACTERICIDAL ASSAY****FACS****ELISA: POSITIVE**

FIGURE 4**PURIFICATION****BACTERICIDAL ASSAY****WESTERN BLOT****FACS****ELISA: POSITIVE**

4/13

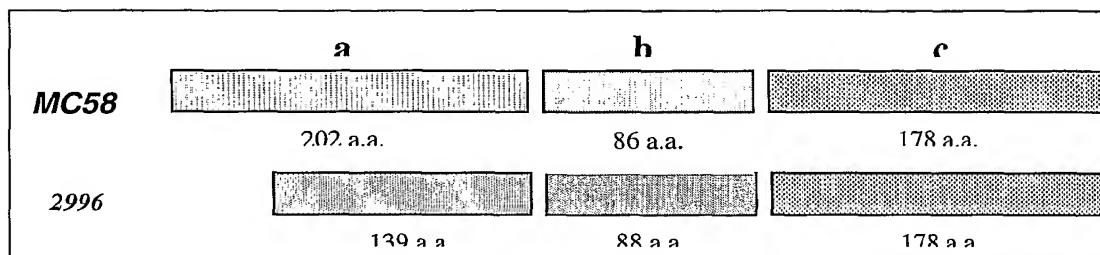
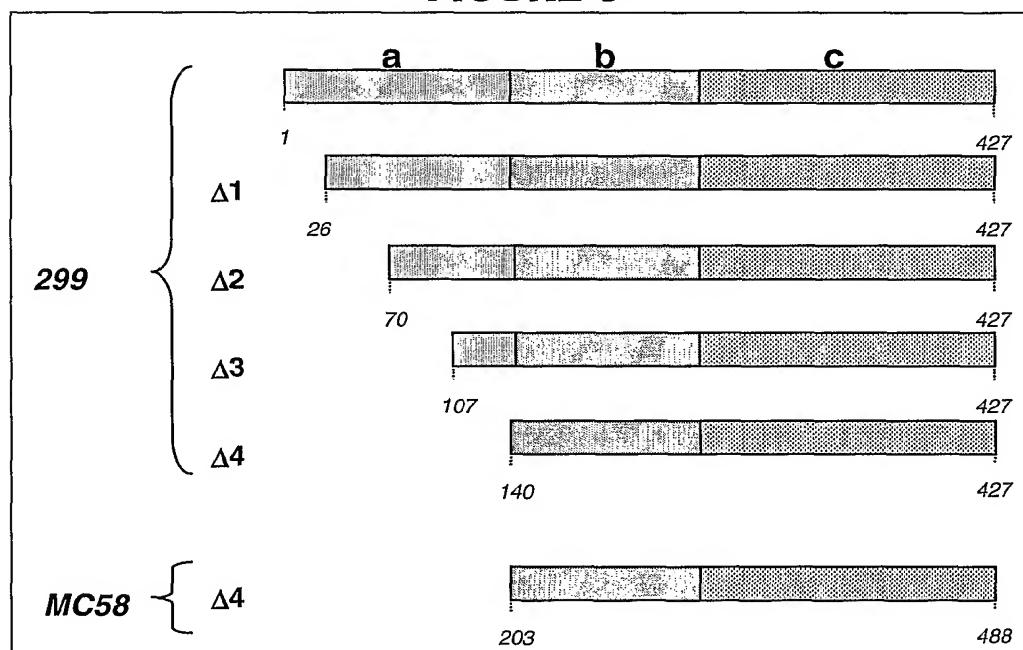
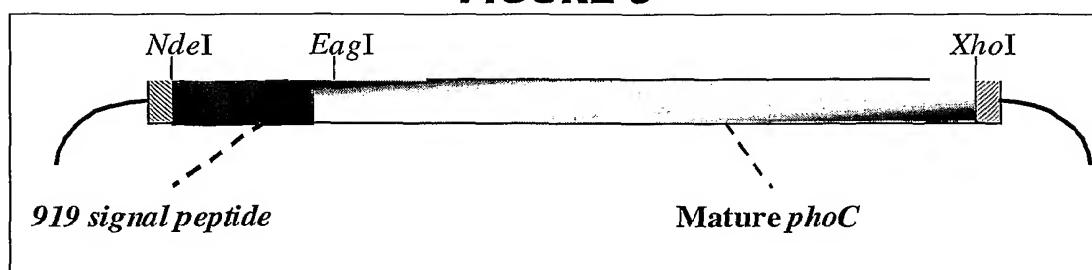
FIGURE 5**FIGURE 6****FIGURE 9**

FIGURE 7

<A> **<Δ1>**

MC58	1	MFKRSVIAACTFALSACGGGGGGSPDVKSADTLSKPAAPVSEKETEAKEDAPQAGSQG
2996	1	MFERSVIAACIFALSACGGGGGGSPDVKSADTLSKPAAPV/AEKETEVKEDAPQAGSQG

<Δ2>

MC58	61	QGAPSAOGSODMAAVSEENTGNGAVTADNPKNEDEVAONDMPONAAGTDSSTPNHTPDP
2996	61	QGAPSTQGSQDMAAVSAENTGNGAATTDKPKNEDEGPONDMPQN.....

<Δ3>

MC58	121	NMLAGNMENQATDAGESSQSPANQPDMANAADGMQGDDPSAGGQNAGNTAAQGANAOAGNNO
2996	106SAESANOTGNNO

-A-><B-

MC58	181	AAGSSDPIPASNPAPANGGSNFGRVDLANGVLIDGPSQNITLTHCKGDSCSGNNFLDEEV
2996	118	PADSSSDSAPIPASNEAPANGGSNFGRVDLANGVLIDGPSQNITLTHCKGDSCNGDNLLDEEA

-B->

MC58	241	QLKSEFEKLSDADKISNYKKDGKNDKFVGLVADSVQMKGINQYIIFYKPK..PTSFAERF
2996	178	PSKSEFENLNESERIEKYKKDGKSDKFTNLVATAVQANGTNKYVIIYKDKSASSSSARFR

<C->

MC58	299	RSARSRRSLPAEMPLIPVNQADTLIVDGEAVSLTGHSGNIAPEGNYRLTYGAEKLPGG
2996	238	RSARSRRSLPAEMPLIPVNQADTLIVDGEAVSLTGHSGNIAPEGNYRLTYGAEKLPGG

<C->

MC58	359	SYALRVQGEPAKGEMLAGAAVYNGEVLFHFTENGRPYPTRGRFAAKVDFGSKSVDGILIDS
2996	298	SYALRVQGEPAKGEMLAGTAVYNGEVLFHFTENGRPYPTRGRFAAKVDFGSKSVDGILIDS

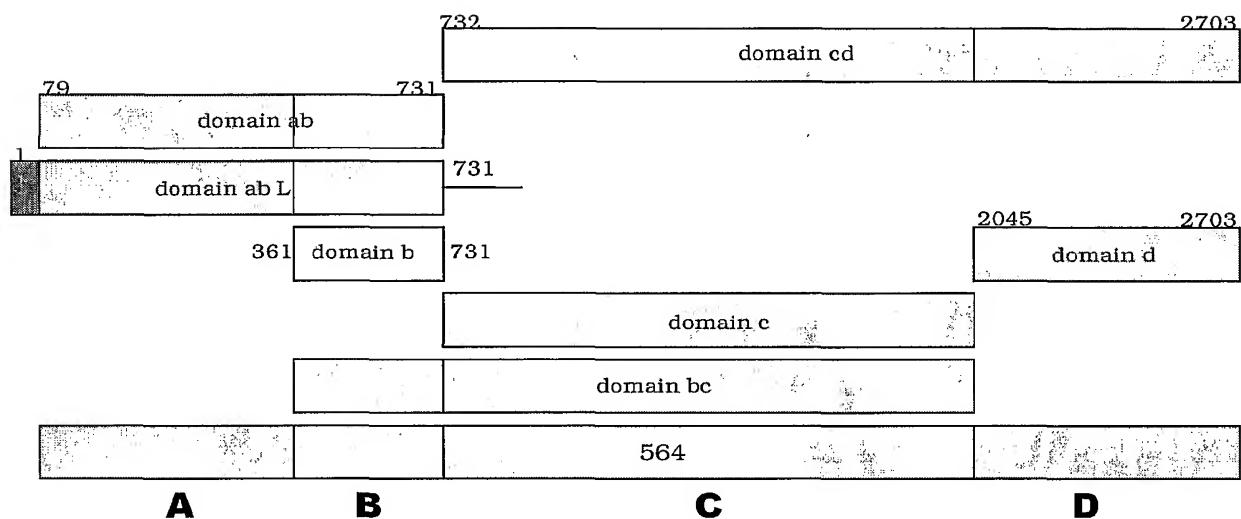
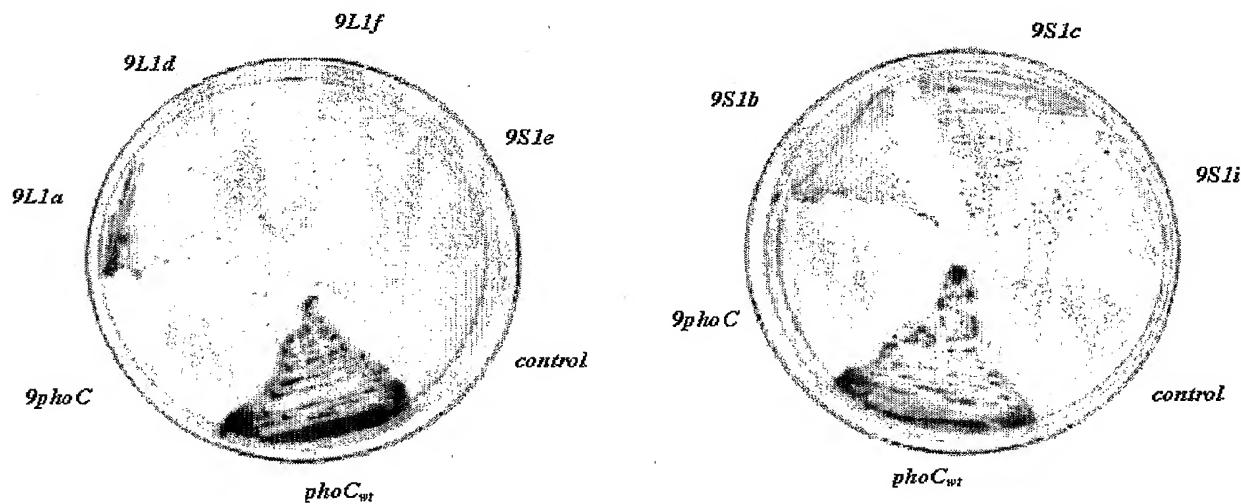
<C->

MC58	419	GDDLHMGTOKEKAIDGNGFKGTWTENGSGDVSGKFYGPAGEEEVAGKYSYRPTDAEKGGF
2996	358	GDDLHMGTOKEKAIDGNGFKGTWTENGSGDVSGKFYGPAGEEEVAGKYSYRPTDAEKGGF

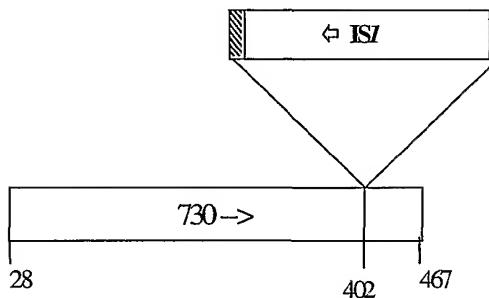
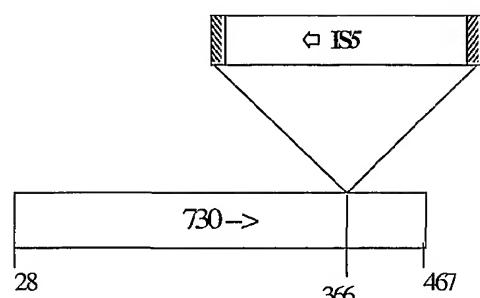
<C->

MC58	479	GVFAGKKEQD*
2996	418	GVFAGKKEQD*

6/13

FIGURE 8**A****B****C****D****FIGURE 10**

7/13

FIGURE 11A**FIGURE 11B****FIGURE 12**

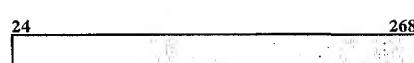
961 (2996)
961 L (2996)



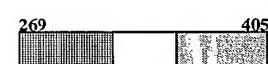
961 (MC58)
961 L (MC58)



961a (2996=MC58)



961b (2996)



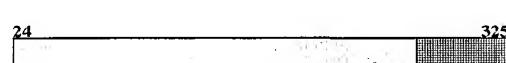
961c (2996)
961c-L (2996)



961c (MC58)
961c-L (MC58)



961d (2996)



961-Δ1 (2996)
961Δ1-L



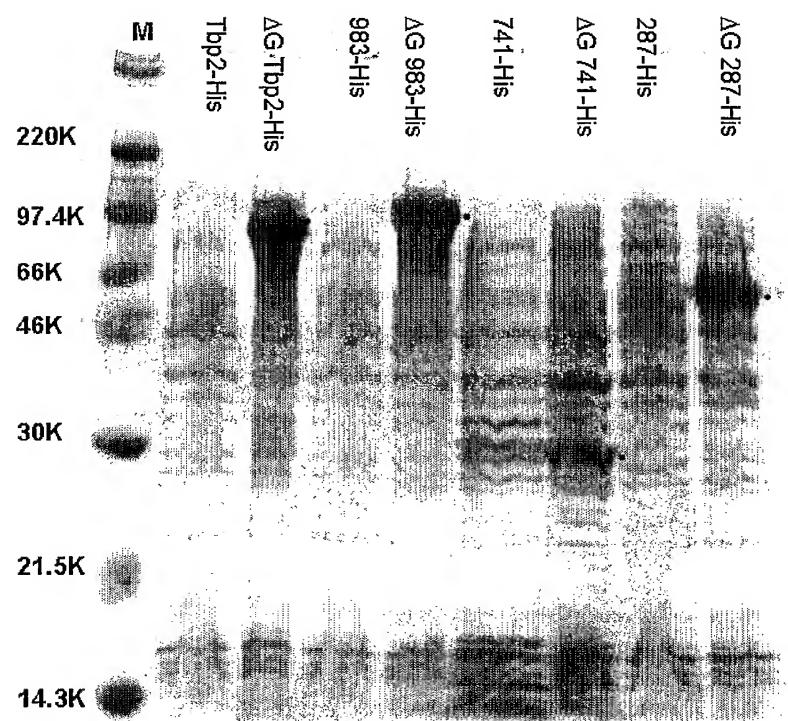
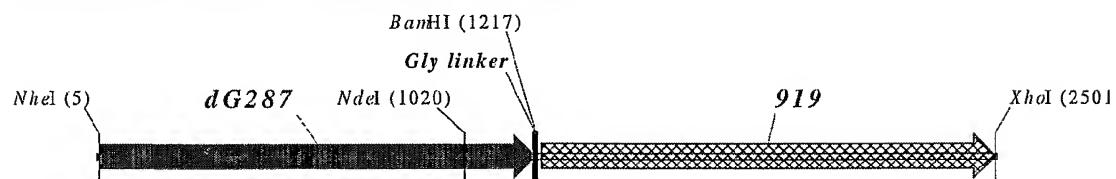
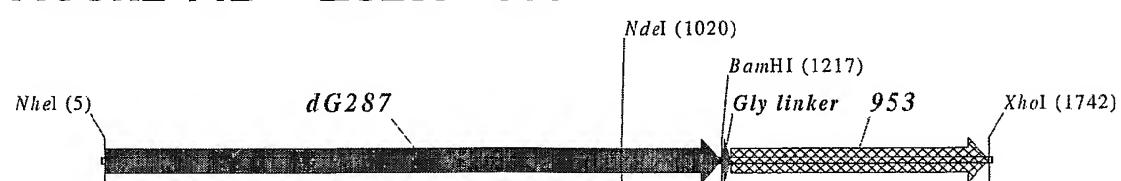
Leader Peptide

Region present in 2996, not in MC58

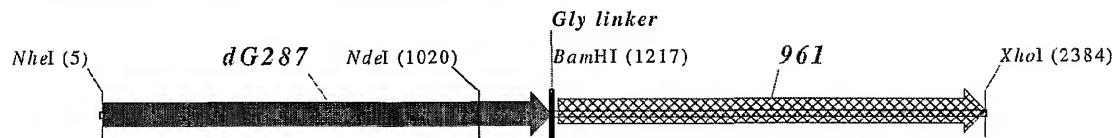
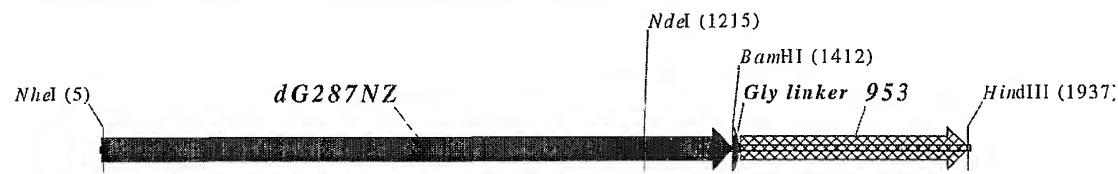
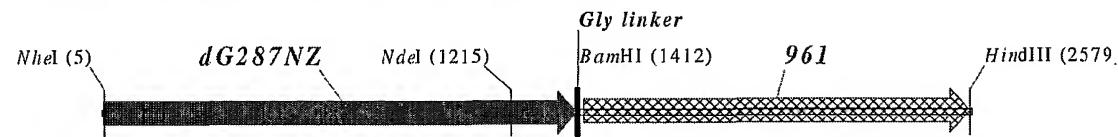
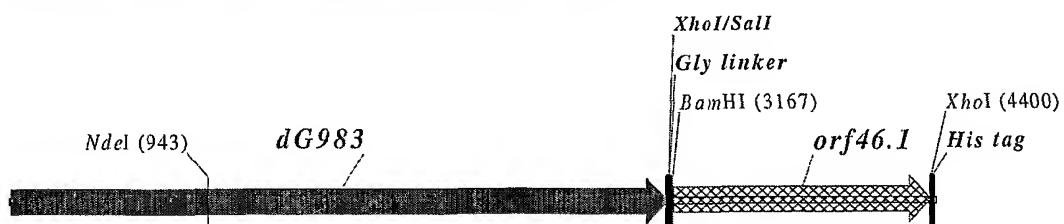
Coil-coiled segment

Membrane anchor

8/13

FIGURE 13**FIGURE 14****FIGURE 14A — ΔG287—919****FIGURE 14B — ΔG287—953**

9/13

FIGURE 14C — Δ G287—961**FIGURE 14D — Δ G287NZ—919****FIGURE 14E — Δ G287NZ—953****FIGURE 14F — Δ G287NZ—961****FIGURE 14G — Δ G983-ORF46.1**

10/13

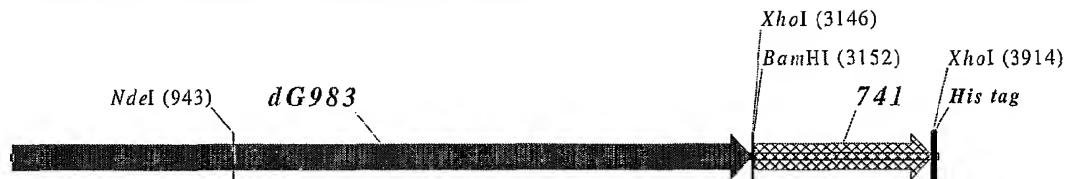
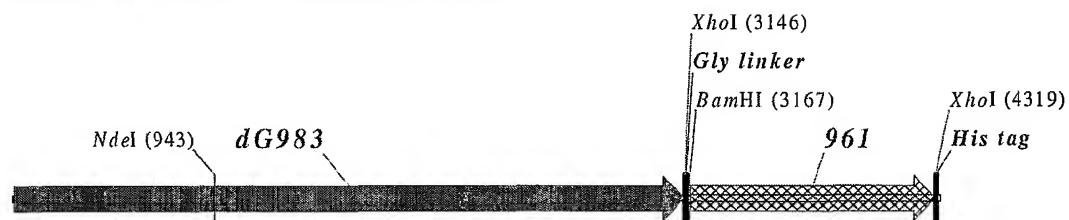
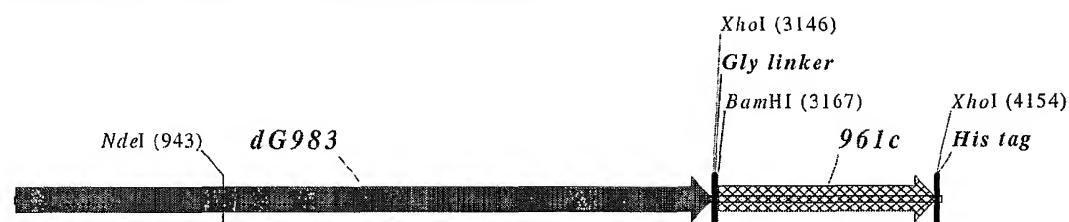
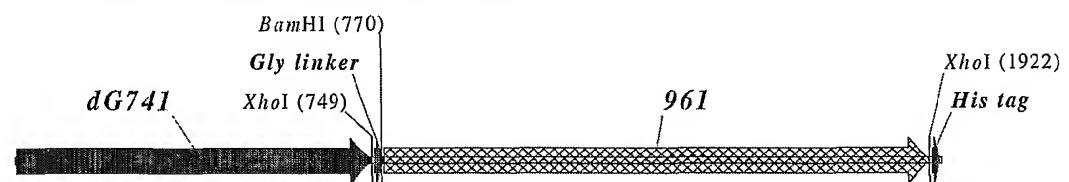
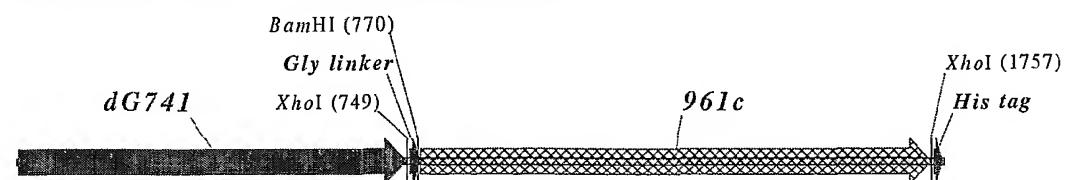
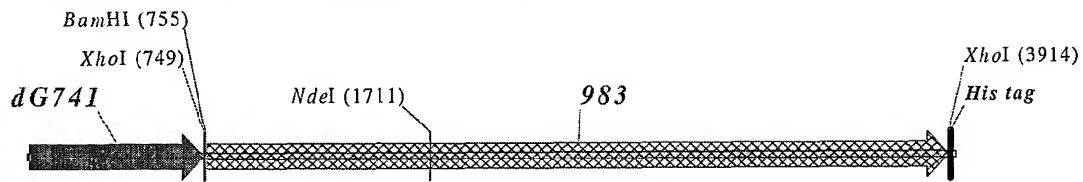
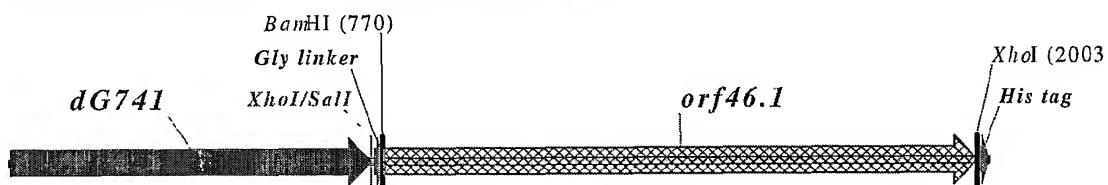
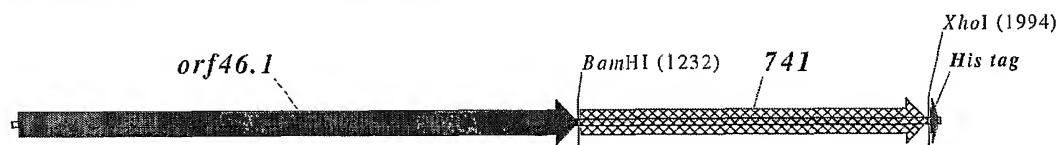
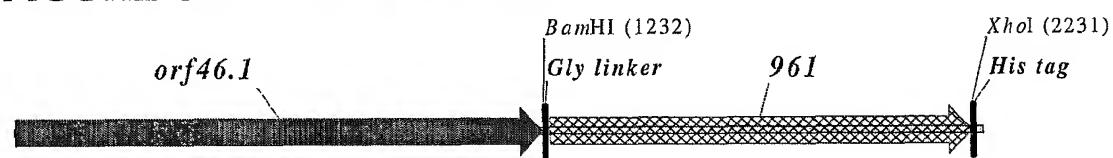
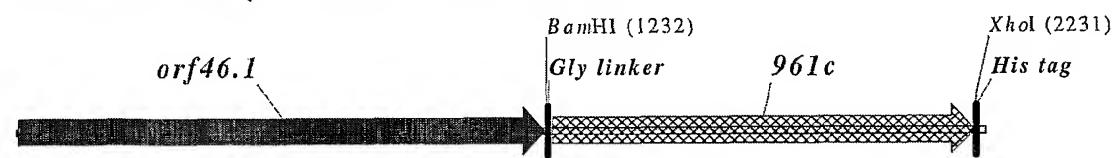
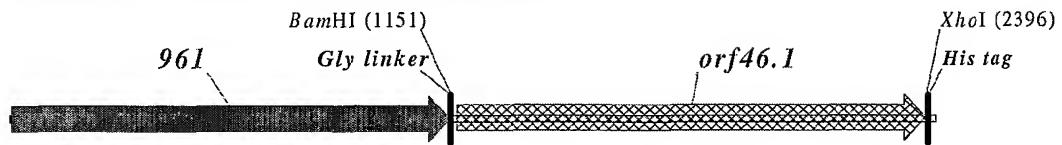
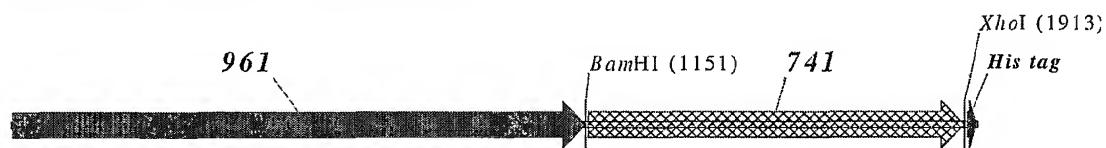
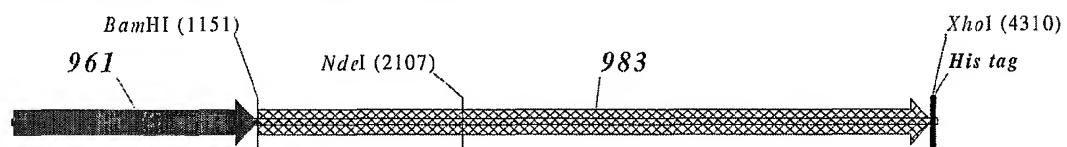
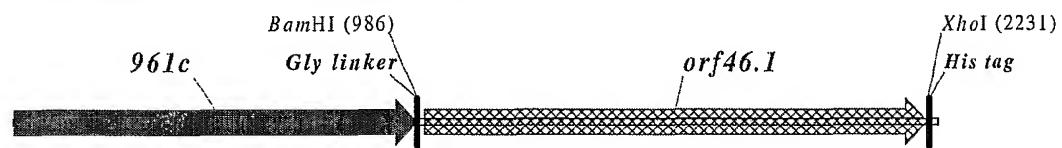
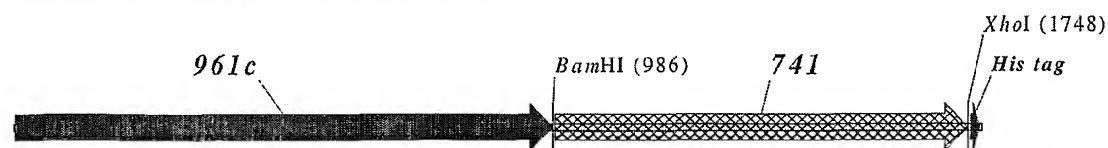
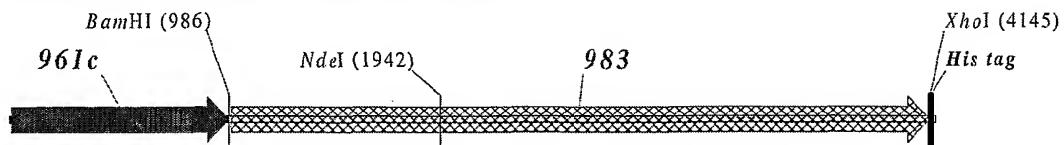
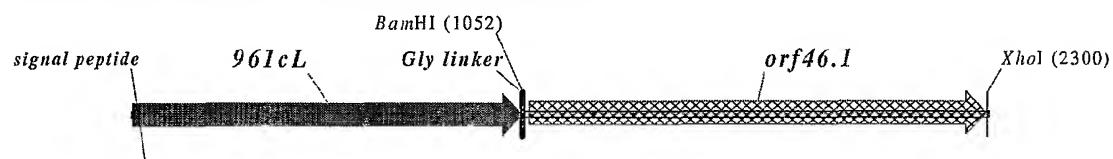
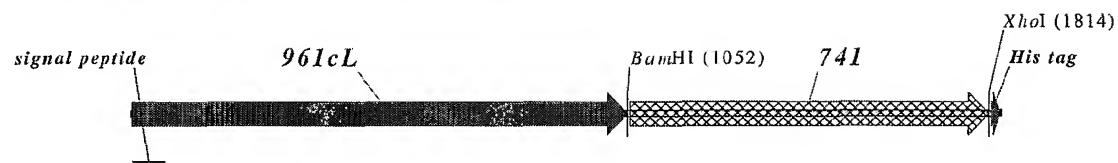
FIGURE 14H — $\Delta G983-741$ **FIGURE 14I — $\Delta G983-961$** **FIGURE 14J — $\Delta G983-961c$** **FIGURE 14K — $\Delta G741-961$** **FIGURE 14L — $\Delta G741-961c$** 

FIGURE 14M — Δ G741-983**FIGURE 14N — Δ G741-ORF46.1****FIGURE 14O — ORF46.1-741****FIGURE 14P — ORF46.1-961****FIGURE 14Q — ORF46.1-961c**

12/13

FIGURE 14R — 961-ORF46.1**FIGURE 14S — 961-741****FIGURE 14T — 961-983****FIGURE 14U — 961c-ORF46.1****FIGURE 14V — 961c-741**

13/13

FIGURE 14W — 961c-983**FIGURE 14X — 961cL-ORF46.1****FIGURE 14Y — 961cL-741****FIGURE 14Z — 961cL-983**